

British science over the hill

The managers of British basic science have gathered ammunition with which to persuade their government towards greater generosity. It may be too late.

BRITAIN is by reputation a stable society, one in which social institutions tend to outlive their usefulness. Thus it was, for example, with fox hunting and the recently abandoned convention that the captain of England's cricket team should not be paid. That is one reason why onlookers are surprised that the British have apparently become indifferent to the wellbeing of the social institutions (universities and publicly supported research institutes) that sustain their research enterprise. But that appears to be the case.

A few decades hence, comparative sociologists will have hit on the obviously plausible explanation, that conservative societies under stress first scrap their newest institutions, on the principle of "last made, first broken". (With the exception of three English universities and three in Scotland, no British university is older than the University of London, now celebrating its 150th anniversary; most research institutes are upstarts by comparison.) Thus will the social history of Britain be tidied into a tale that can be taught to students. By then, it will seem entirely irrelevant that, in 1986, the British government's own advisory committee on strategy for basic science (the Advisory Board for the Research Councils, or ABRC) commissioned two reports that catalogue the decline of British basic science and that, by having what Sir Peter Swinnerton-Dyer calls "sharp corners" (see page 659), may be used to persuade the British government to enlightenment or at least generosity. The stragem will be discovered not to have worked because those who devised it failed to understand the nature of the attack upon them.

None of this detracts from the value of the two documents published last week by ABRC; a summary of one appeared last week (see *Nature* 323, 591; 1986) while a summary of the second can be found on page 681 of this issue. The former study amounts to a judicious definition and assessment of what Britain and comparable countries spend on basic science. To students of the field, one of its chief merits will be the way in which the data on research spending serve to correct the often false impressions given by international comparisons based on nationally provided information; this is a lasting public service. Like other exercises of this kind, the study stops short in 1982, which is when the iron began to bite in Britain. One conclusion is that British spending on basic research between 1975 and 1982 grew less quickly than in the United States, France and Japan (but in West Germany there seems actually to have been a decline over the same period). That is qualitatively well known. The most striking lesson to be learned from this document is that public spending on research in Britain is absolutely low, which is not merely a consequence of government policy but of Britain's declining economic fortunes.

Citations

The second report, by the Science and Technology Policy Studies Unit of the Royal Society to which has recently been conjoined the ambiguous sponsorship of the Fellowship of Engineering, hangs on bibliometric analysis, or an analysis of citation counts. The Philadelphia-based Institute of Scientific Information, which generates the basic citation counts, will no doubt

be grateful for this careful demonstration of the utility of its data, as refined under the auspices of the US National Science Foundation. Dispassionate readers will note from this interesting compilation that the British performance in research is still reasonably respectable; the volume of publication in any field is almost always less than that from the United States, but not too different from countries such as France, West Germany and, sometimes, Japan. But, ominously, the trend is one of decline even in the fields (such as genetics) in which the data show Britain to have been strong. This again accords with anecdotal impressions of what is happening in Britain; desuetude rather than catastrophe is the most likely outcome.

The managers of the British scientific enterprise say they plan to use the publication of these two reports to tell the British Treasury that the former describes a cause (underspending) and the latter its effect (declining output). The argument will be lost before it can be made; the hard men will simply say that this relationship between cause and effect is what they had always intended.

Budgets

Presumably to its credit, the present British government has not hidden its intentions for the research enterprise. Voluntarily (Mrs Margaret Thatcher in 1980), it undertook to "protect" the aggregate budget of the grant-making research councils; arbitrarily, but explicitly, it told the universities (Mr Mark Carlisle, December of the same year) that university budgets must be cut by at least 8.5 per cent (which turned out to be a cut in real terms of about 14 per cent after allowing for the regulation of overseas students' fees); it has also pleaded with anybody who would listen that the research enterprise in Britain should be more conscious of the needs of British industry. The policy has been open and consistent. (That it may be mistaken is, of course, a different matter.) Moreover, on the evidence assembled by ABRC, this policy has been successful; does not the decline of the number of influential research articles betoken the evacuation of the hated ivory tower? ABRC's difficulty, at the meetings that lie ahead, will be to explain why it is such a bad thing that the emphasis should be shifting from basic to applied research when the total volume of research activity is, at best, stagnant.

Quickly (if its intentions are serious), ABRC should learn by heart the following precepts. **Enforced rapid change at times of static budgets can be uneconomic.** Many of the research councils have been forced within their protected budgets to retrench, wasting money and people in the process, but there has been a greater waste in the rapid reduction of university staffs, especially as many departments have been squeezed below their economic size. **Not all basic research is academic self-indulgence.** The development of many of the fields of applied science, such as the design and use of microelectronic devices, in which the British government would give its eye-teeth for signs of progress, are at present frustrated by the lack in Britain of people familiar with new techniques and by the patchy coverage of the basic physics by university and other laboratories. **It serves no purpose to persuade basic researchers into applied research if the products**



of that activity are overlooked. British industry has an unenviable record for looking innovatory gift-horses in the mouth. **Applied research, being part way towards development, is more labour-intensive than basic research.** Has anybody in Britain calculated how little will be accomplished by re-treaded academics? **The morale of the scientific community matters more than its organization or budget.** And, in Britain, morale is near rock-bottom.

The next few months will tell whether such arguments, sustained by the necessarily dated evidence of ABRC's two reports, will carry weight with a government that, so far, has been indifferent to them. Some sign of which way the wind is blowing should come next month, when estimates of public spending for the next three years will be published. The chances are not bright. All the arguments have been made before. And, given the unavoidable delay in the compilation of statistics on the pattern of research spending and achievement, there is little that ABRC can do in the time available to prove the most chilling assertion about the condition of British science: it is so long since a bright person with a bright idea could reckon on finding the support needed for its pursuit that creative people know their future lies outside Britain (where good work is, in any case, better rewarded).

In the long run, that is how the enterprise will be killed off. The most skilled will leave, leaving the conduct of research and the education of the young to those, not an unbiased sample, who happen not to have done so. Meanwhile, the institutions most at risk, the universities and the research institutes, will most probably compliantly wait for the reforms that could give them a better chance on the leisurely proceedings of complicated committee networks certain to yield compromised half-reforms. The planned change in the transition between secondary and higher education in Britain is a case in point. After twenty years in which far-sighted universities have conspired with short-sighted governments to preserve the uniquely specialized character of secondary education, it is now planned that there should be a marginally broader curriculum leading from school to higher education two years from now. In a decade or so, entrants to the research professions may be marginally more broadly educated than has been the rule, although still more blinkered than in most other places. But why wait ten years for an unsatisfactory solution of a long-standing problem? Why, in the field of public administration in which ABRC is charged with responsibility, wait for the benefits of the slow processes by which UGC plans to concentrate research activity on the more productive universities? Could not those in charge find a quicker way? Or find others who might take the risk? The trouble with these conservative societies is that they require that institutions under pressure should be compliant in their own decline. The present management has commissioned two useful reports. Might it not also speak out and say what should now be done? The British government might not listen, but the troops would be modestly encouraged. □

UNESCO's new future?

Member governments of the UN's chief cultural agency should seize the opportunity for reform.

THERE are two views of UNESCO, the United Nations Educational, Scientific and Cultural Organisation. One is that its constitution is so vague and its ambitions so far-reaching that very little should ever have been expected of it. Another is that it is a splendid organization which, by bad luck, has recently fallen on hard times, sacrificing in the process the membership of the United States, Britain and Singapore. The truth most probably lies somewhere in between. There may also be some who hold that UNESCO is a splendid organization which remains in splendid condition, but little has been heard from them in the recent past. Whichever may be the correct opinion, the mem-

bers of UNESCO will have an opportunity in the next few months to edge the organization in a direction better suited to an intelligent reading of its purpose.

The immediate occasion is that the director-general for the past ten years, Mr Amadou-Mahtar M'Bow, is coming to the end of his second term in office. Already, the executive board in Paris is searching for possible successors, not necessarily incommoded by the position on the issue of Mr M'Bow, who has said that he will not seek a third term but will not decline to serve if asked. Mr M'Bow personally has been at the bottom (or top) of many of UNESCO's troubles in the past troubled decade. He appears to have brought to the organization a style of management that left simple servants of UNESCO in perpetual uncertainty about their functions or even the continuation of their jobs, although there may be nothing more to this than the way in which vain people often demand unsustainable loyalty from their colleagues. Mr M'Bow's more real failing is that he has done so little that can be called new. He was personally committed to the scheme for creating a "new world information order" which, whatever its rights and wrongs (the second predominated), might have been calculated to drive away the United States at least; curiously, with the departure of the United States two years ago, very little has been heard of this strange proposal. It would be in the organization's best interests that Mr M'Bow should on this occasion be taken at his word, and allowed peacefully to pursue other interests. It would be disastrous if he were to stay.

The issue of who shall be director-general is, unfortunately, only a small part of UNESCO's problem and, most probably, more a symptom than a cause. Since its foundation in the heady days at the end of the Second World War, UNESCO seems profoundly to have lost a sense of what it is about. Although the organization's budget has never been as big as that of a dozen or so private foundations in the United States, its members have allowed it to interpret its field of action (and competence) in the widest way, and have urged it to become a kind of impoverished aid agency as well.

The organization is also perpetually undecided about the function of its staff. Is UNESCO a grant-making agency, or a kind of ivory tower whose occupants have the wisdom and the practical experience to solve problems at which others have failed? Disastrously, UNESCO has from the outset nursed the second conceit. What its members should now seek is a formula that would allow the human beings who work loyally for it to do work that is within their competence, a quality that itself has been eroded after several years of quarrelling. Can that be too much to ask?

Unhappily for some of UNESCO's well-wishers, a move in that direction would not necessarily bring lapsed members rushing back to the fold. A large part of the reason why the United States and, later, Britain left was their belief that UNESCO had become "politicized". To the extent that that amounts to a protest against the habit of groups of UNESCO members of ganging up against each other, the charge is valid. But disaffected Western governments have too often in the past complained about UNESCO simply because many of the issues that crop up are intrinsically political. It is, for example, squarely within UNESCO's terms of reference to worry about the preservation of important archaeological sites; doing that diligently must step on sovereign toes. The same is true of other important issues that UNESCO has successfully negotiated, copyright for example. Even the new world information order was a proper subject for UNESCO to take up; the fault was merely that an undistinguished director-general allowed a biased proposal to take the centre of the stage. A wiser person would not necessarily have won friends among those who took offence, but might have accomplished something. The lesson is that Britain and the United States should not pretend that re-entry would be a bed of roses. UNESCO membership will always need a strong stomach. □

US budget

Congress scrambles last-minute omelette

Washington

CRUCIAL decisions on arms control, science budgets, a replacement space shuttle and environmental clean-up were all crammed into the days before the final gavel ended the 99th Congress last Saturday. After first performing what is now the annual ritual of letting the government run out of money, Congress passed a \$576,000 million spending bill and raced off for two weeks of campaigning before next month's national elections.

The major spending bill, known as a continuing resolution, received final approval on Saturday. The legislation, containing virtually all appropriations for the entire federal government, will require some time to decipher. The National Institutes of Health received more than \$6,100 million, more than the amount requested by the Reagan administration. The National Science Foundation (NSF) escaped cuts proposed by the House of Representatives, ending with an appropriation of \$1,622.9 million, approximately \$63 million less than had been requested. Congress said that NSF should spare programmes in global geoscience, ocean studies, astronomy and minority programmes from the now necessary cuts.

The National Aeronautics and Space

Administration (NASA) budget, also determined late in the game, includes \$2,400 million for a fourth shuttle orbiter. This money is extra to other NASA programmes, not in place of them as many at the agency had feared. The Department of Energy budget allows research on the Superconducting Supercollider to continue, but reiterates the demand that the Energy Department submit specific plans on how it plans to finance the project.

Congress and President Reagan nearly reached an impasse over the continuing resolution when the president objected to language on arms control included in the proposed legislation (see *Nature* 322, 765; 1986). Democrats in the House of Representatives sought to force the president to halt production of chemical weapons, abide by the terms of the SALT II treaty and refrain from further nuclear tests so long as the Soviet Union did likewise.

Under the final compromise, the Pentagon may proceed with procurement of 155-mm shells containing binary nerve gas, but the procurement of the much criticized Bigeye bomb is delayed one year (although \$90 million was released for a facility to build the bomb). On the nuclear testing issue, Congress traded its insistence on a moratorium for an agreement

from the President to submit to the Senate for ratification the Threshold Test Ban Treaty (signed in 1974) and the Peaceful Nuclear Explosions Treaty (signed in 1976). The president has also vowed to work towards a comprehensive test-ban treaty. The bill also contains a strong but not binding form of words urging the president to abide by the terms of SALT II.

Money for the Strategic Defense Initiative (SDI) was held to \$3,530 million, well below the \$5,400 million the administration was seeking. A proposal that caused great consternation in Europe to limit contract SDI work to US companies failed to survive into law.

Among the bills that foundered at the end was the extension to the Price-Anderson Act, which originally established the limits of liability for nuclear power plants in the event of an accident. Under the present law, each operating plant would be liable for up to only \$5 million per accident, an amount generally reckoned insufficient to cover the cost of a major accident. The proposed extension would have raised the liability limit to \$63 million per accident per reactor. The legislation died when a proposed compromise failed to coalesce support from both environmental groups and the nuclear power industry necessary for the bill to pass.

A bill restoring to life the moribund Superfund toxic waste clean-up programme was signed into law last week. The president had threatened to veto the bill, objecting to the cost of \$9,000 million over 5 years allowed by Congress as well as to the broadly based corporate tax enacted to pay for the programme. Congress also threw together a \$1,700 million omnibus drug bill setting tougher penalties for trafficking in illegal drugs, establishing programmes for crop eradication and substitution and providing new funds for education and rehabilitation.

A major effort to redefine steps for compensating victims of childhood vaccine-related injuries achieved a half-way existence. Congress passed certain changes to the legal system for pursuing claims, but failed to enact an excise tax on vaccines that would pay for an insurance pool to cover claims. The legislation will have no effect until measures are passed to raise the revenue to pay for it. Both the Department of Health and Human Services and the Justice Department oppose the legislation, and President Reagan may veto it.

Allowing the government to run out of money, as happened twice before stop-gap measures could be passed, does not endanger critical government functions, but can have some odd consequences. During one such funding gap last Friday, a call to the White House Press Office was answered with the curt statement, "I'm sorry, the White House is closed".

Joseph Palca

Unapproved drugs for export

Washington

IN the pandemonium of its last few hours, the 99th Congress finally passed a long-fought measure to allow US pharmaceutical and biotechnology companies to export drugs and biologicals before they have been approved for use in the United States. Foreign importers will have to certify that products will not be re-exported. If signed into law by the President, the measure would be a boon for the industries. But the measure was included in a package of health measures that includes controversial legislation to compensate childhood victims of reactions to vaccines and could yet draw a presidential veto.

The biotechnology industry has found itself hampered by the present laws forbidding the export not only of drugs but of any biologically active substances not approved for domestic use. The new act specifies 21 countries to which exports of drugs and biologicals will be permitted because they are deemed to have adequate regulatory systems of their own.

Companies wanting to export unapproved drugs must still request permission

from the Food and Drug Administration to do so, giving 90 days' notice. And the drug in question must already have received an "investigational new drug" permit, required before clinical trials.

The drug export legislation is opposed by some who believe it represents a double standard, allowing drugs to be used abroad that have failed to gain approval at home. But manufacturers point out that no other countries have similar export restrictions and that many countries approve drugs more quickly than FDA.

The legislation was held up in Congress for several months by Representative Henry Waxman (Democrat, California), who wanted to condition his approval on the Senate's accepting a congressional study of pharmaceutical labelling practices in developing countries. Many pharmaceutical companies reportedly omit warning information on products for the domestic market. But Waxman eventually agreed to the drug export provision when it became clear that some of his favoured health legislation, including vaccine compensation, would otherwise die. Tim Beardsley

British science politics

Better prospects at polls?

THE mechanics of transferring the results of scientific research, principally defence work, from the laboratory to the shop floor has become a political issue in Britain and could soon result in the axing of some projects.

The government will publish next month the results of a major study of public support for scientific research, while the Labour party, the principal opposition party, has promised that, if it wins the next general election, it will create a Ministry of Science to ensure the better exploitation of research.

The new study, which has taken about a year to complete, is an attempt by the government to assess whether it is "getting value for money" and a satisfactory return on the £4,600 million now invested each year in research and development.

The government has long been convinced that there is an inadequate bridge between the research institutes and the commercial sector. The review, supervised by an unnamed cabinet minister, has received advice from the new watchdog on government supported research, the Science and Technology Assessment Office, a unit within the Cabinet Office.

The study, covering all government department's spending on research and development, is likely to recommend a reallocation of resources; changes are likely, in particular, at the Ministry of Defence and the Department of Education and Science.

The Labour party last week confirmed its intention to create a new ministry, "manned by people who understand science", to ensure the commercial exploitation of the findings of government-supported research; to that end, the ministry would have the power to breach traditional departmental boundaries.

The third major political party in Britain, the Social Democratic Party Liberal Alliance, is equally dissatisfied with technology transfer from academic institutions and government-supported research programmes to the assembly line. Its approach would be to create a structure outside government departments with a £200 million launch budget and simultaneously to increase support for research by about £40 million a year.

The Alliance policy endorses the view held by the other two parties. Its policy document says that although government must provide the framework within which industry may develop in a competitive environment, the needs of the new technologies are so urgent that Britain "cannot wait for the invisible hand of the market to take effect"

Bill Johnstone

Soviet Union

New president for academy

THE Soviet Academy of Sciences has a new president, an appointment that seems to conform to Mr Gorbachev's policy of breaking down inter-ministerial barriers in science and technology. Academician Gurii Marchuk, former director of the Siberian branch of the Academy of Sciences of the USSR (1975-80) and more recently head of the State Committee for Science and Technology (1980-85), was last week unanimously elected to the office relinquished by the 83-year-old Anatolii Aleksandrov who is said to have retired "at his own request". Marchuk was elected by secret ballot, but as with Aleksandrov, and unlike earlier presidents, the proposing speech was made not by a scientist, but by Politburo Member and Central Committee Secretary Yuri Ligachev.

Marchuk carries on the tradition of the past several decades, which have entrusted the leadership of the academy to mathematicians or physicists. The previous two presidents (excluding the caretaker president Vladimir Kotel'nikov) have been an expert in rocket technology (Mstislav Keldysh) and a nuclear physicist (Anatolii Aleksandrov), whose appointments largely coincided with the prevailing priority in Soviet research when they were elected.

On that basis, it would be difficult to identify the current Soviet priority, for Marchuk is an all-rounder. After graduating from Leningrad University, he became a professor of mathematics, specializing in the dynamics of large-scale atmospheric processes, and then a researcher at the Geophysics Institute of the Academy of Sciences. He then moved on to nuclear power engineering, and then to computer technology, founding the Computer Centre of the Siberian Branch of the Academy of Sciences. Members of the academy last week paid tribute to his democratic outlook and flexibility as an administrator, and noted that during his time at the Siberian Branch of the academy, virtually all discoveries made there were implemented in Soviet industry. According to Boris Paton, president of the Ukrainian Academy of Sciences, Marchuk "shows an equally good knowledge of fundamental science and industry".

But in proposing Marchuk for the presidency, Ligachev did not confine himself to eulogy of his candidate. He used the occasion to castigate the academy for a loss of initiative in the long-term planning of promising fields of science, and for paying insufficient attention to "general state programmes". This system, Ligachev warned, can and must be rectified.

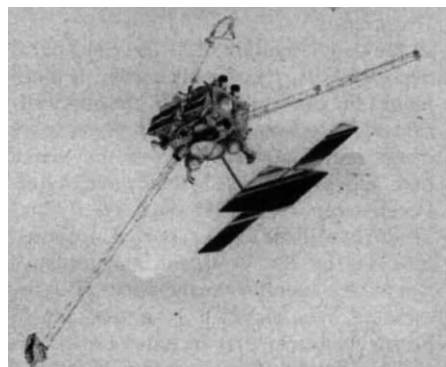
The Central Committee of the Communist Party of the Soviet Union is pre-

pared, he noted, to consider the problems of development of the academy, to define measures to resolve pressing questions, and, in particular, to look at research teams that are not sufficiently productive. The role of regional branches of the all-Union academy and the fourteen academies of the union republics must be improved, said Ligachev, to improve the "labour efficiency" of every research team in the country.

Replying to these criticisms on behalf of the all-Union academy and its union-republic associates, Marchuk repeated Ligachev's calls for efficiency, and called on all scientists to rethink the role of science at this particular stage of Soviet development. He called for the branches of the academy to be transformed into "headquarters which elaborate the strategy of science and determine scientific priorities and the direction of research" and, in particular, pledged his support Mr Gorbachev's science policy. **Vera Rich**

Mars spacecraft

IF the satellite picture looks more like a communications satellite than a planetary probe, that is not surprising. Last month, the National Aeronautics and Space Administration (NASA) signed a contract with RCA Astro-Electronics Division to build the Mars Observer spacecraft, the first of what could be a series of relatively low cost planetary missions drawing heavily on existing satellite technology. The Mars Observer structure is based on



RCA's Satcom communication satellite. The spacecraft will carry eight instruments to study the magnetic and gravitational fields, the physical characteristics of surface material, and the atmosphere and seasonal changes of Mars. Project manager William Purdy says the point of the observer series is to spend "more money on the science than on the spacecraft". The budget for the mission is approximately \$252 million, and the spacecraft costs about \$50 million. No specific launch date has been set, but 1990 is the target. □

US innovation

Technology transfer on the move

Washington

SWEEPING changes designed to speed the commercial uptake of technology from US federal laboratories are expected to be signed into law this week by President Reagan as the culmination of a six-year effort by Congress to liberalize private-sector participation in federal research. Despite some earlier agency opposition, all 700-odd federally operated laboratories will in future be encouraged to carry out cooperative research with private companies, universities and state organizations, even assigning patent rights to collaborators when that will accelerate transfer. The stimulus for the new act was the perception that foreign competitors are making better use of US government research than is domestic industry.

More controversially, the Federal Technology Transfer Act of 1985 also obliges federal agencies to reward employee-inventors with not less than 15 per cent of patent royalties arising from their work, up to \$100,000 a year. Much of the balance will be set aside for education and incentive schemes for others who make important contributions.

By promising federal employees better personal gains for stimulating technology transfer, and by offering private industry the carrot of title to inventions, the new act provides incentives for both technology 'push' and 'pull'. It stipulates that laboratory managers should give preference to domestic industries and to small businesses.

The royalty-sharing provisions are nevertheless opposed by some of those with private interests in intellectual property. The Intellectual Property Owners' Association, a lobbying group for large research corporations, argues that such provisions unreasonably limit the discretion of managers to reward as they see fit. It also fears the law may encourage private industry to adopt the same rules, a prospect it views with horror.

The research pool to which US industry has now been given the keys is considerable. Federal laboratories employ one out every six scientists and engineers in the country, and consume an annual budget of \$18,000 million. Hitherto, the rules governing collaboration have been a patchwork, with only about a half of federal laboratories legally allowed to enter cooperative schemes. Some agencies, such as the Department of Energy and the National Aeronautics and Space Administration, already have cooperative projects; others, including the Departments of Defense and Agriculture, have argued that they lack the legal authority to do so.

The new law puts all agencies on the same footing, and extends legislation pas-

sed in 1980 and in 1984 that started the technology transfer ball rolling. The big contractor-operated laboratories run by the Department of Energy, such as Argonne and Los Alamos, were given formal authority to engage in collaborative research two years ago. Henceforth, all federal laboratories, whether contractor- or government-operated, will be able to contribute and to receive resources as part of cooperative research agreements.

Another provision of the new law formally recognizes the Federal Consortium for Technology Transfer, which is intended to stimulate uptake of technology by providing training for laboratory staff and information for industry. An informal

predecessor organization has existed for more than a decade, but with a budget of only \$200,000 per year its influence has been limited. Its new incarnation has been given a permanent home in the National Bureau of Standards and \$900,000 to set up an electronic mail system linking the laboratories.

The new rules raise some awkward questions about the federal role. Brian Frost, director of technology transfer at Argonne National Laboratory, says laboratories will "have to walk a careful line" to ensure that federal funds are not used to support research that should properly be undertaken by industry. Similarly, nobody wants industry money to substitute for federal funds. But if correct decisions are made, Frost believes the result will save the taxpayer money while improving uptake.

Tim Beardsley

British research

Dispirited researchers on ropes

BBRITISH researchers last week responded despondently to the publication of two reports on the state of basic science commissioned by the Advisory Board for the Research Councils (ABRC). At a meeting in London to discuss the documents (see page 655), the leaders of the research and academic communities were more dispirited than hopeful.

That "something modestly good" would come out of the publication of the two reports was the most that Sir David Phillips, chairman of ABRC, and Sir Peter

reports shows that "cheap methods work"; he would like to see the study extended from genetics and physics to all branches of British science. But he also warned against assuming that, if a field is shown to be weak or weakening, it should therefore be strengthened. Britain is still getting better value for its basic research spending than any other country, he says.

In Britain, the research councils and UGC should remain responsive, not directive. "I would rather back the aggregate of decisions by scientists in their twenties than the views of ageing committees, which are good only at deciding what should have been done 30 years ago."

On UGC's still-controversial procedures for deciding how to concentrate research in particular universities, Swinnerton-Dyer said that bibliometric indices would have no place in short-run assessment, that the valuable device of engaging experts from abroad in the assessment of research projects would require a much larger administrative budget than UGC now has, that UGC's present network of committees had the merit of being quick and flexible and that peer review would remain paramount in its decision-making.

ABRC may be able to use the hard data in the documents for future arguments with the Treasury. But Phillips also emphasized the necessity of choice: if Britain is responsible for roughly 5 per cent of the world's science, should it be in every field or should it simply back a few leading individuals, setting them up in larger units?

Britain's real problem, Phillips said, is that the rapid growth of the period 1966-73 had been followed by static budgets at a time when world scientific spending had doubled in a decade. "We can see enormous opportunities in Britain, but they cannot be met".

Robert Walgate



Swinnerton-Dyer: good value for money?

Swinnerton-Dyer, chairman of the University Grants Committee (UGC), could offer. Indeed, they seemed more ready to talk of further rationalization (newspack for cuts and economies) than of strategies for raising support for British science. Nevertheless, Swinnerton-Dyer said the "sharp corners" of the reports would help in dealing with the Treasury.

Swinnerton-Dyer said that the concurrence of the conclusions of the ABRC

Genome sequencing

More actors apply for parts

Bethesda, Maryland

THE US National Institutes of Health (NIH) last week sponsored the latest instalment of the human genome mapping and/or sequencing debate, and once again no clear winners emerged. But there has been progress. The initial polarization of opinions has given way to a more constructive consensus that some concerted effort can begin without rending the fabric of biological science.

Like many other institutions, both public and private, NIH are now grappling with the question of what their role should be in any sequencing or mapping project. Last week's discussions were held at the meeting of the advisory committee to NIH director James Wyngaarden. Given budgetary lead times, any major financial commitment from NIH could not come before the 1989 fiscal year.

At the heart of the debate is the question of the value of knowing the sequence of the entire genome. David Baltimore of the Whitehead Institute argues that a free gift of the sequence would be welcome, but not having it was no great hindrance. If asked to rate the project, he would place it in the category of worthy projects that could be funded if unlimited resources were available. As that is not the case, Baltimore would prefer to see funds spent on smaller projects with more immediate promise of interesting results. Baltimore believes that a huge, low-priority project would be "ruinous" and could easily become a political target. David Botstein of the Massachusetts Institute of Technology

argues that a "pay-as-you-go" approach would provide sequence data in a piecemeal fashion, but the emphasis would be on the 3 per cent of the genome thought to encode protein. Botstein questions the need "to sequence everything to get to the 3 per cent". But Renato Dulbecco of the Salk Institute believes that only a complete sequence will reveal the functional significance of how the chromosome is organized.

If pursuing a sequence remains a contentious area, there is fairly wide support for making both a genetic and a physical map of the genome. Most concede that a complete physical map ought to precede any full-scale sequencing. But efforts so far in both sequencing and mapping have shown a need for better data handling and organization procedures. Wyngaarden believes that information organization and communication is the area most urgently needing attention even if sequencing and mapping efforts never get any bigger than they are now. Work so far is proving nearly overwhelming for GenBank, the DNA sequence database supported by the Department of Energy (DoE) and NIH.

George Cahill of the Howard Hughes Medical Institute told the NIH meeting that coordinating emerging databases was rated a top priority by participants in meeting on the human genome held at his institute last July (see *Nature* 322, 397; 1986). Hughes is currently supporting work by Frank Ruddle at Yale University on a human gene map database, and

hopes to link that database with Victor McKusick's Mendelian Inheritance in Man recently available on line. Next year, Cahill hopes to see Jackson Laboratory's mouse gene data linked with the human databases.

The need for a well-organized information exchange system is felt keenly in Europe. Lennart Philipson, director-general of the European Molecular Biology Laboratory (EMBL), argues that a forward thinking plan is needed to make full use of the information that will be pouring in from laboratories in Europe, Japan and the United States. That such an exchange system is possible is demonstrated by EMBL's participation in GenBank. EMBL and NIH will hold a joint meeting to discuss standardization next February.

Besides storing sequence and map data, there is a need to make sense of it. Both the National Science Foundation and DoE are supporting efforts to develop new algorithms for handling such data.

Although NIH seems a logical government agency to be interested in the human genome, other agencies such as the National Science Foundation, DoE and even the Department of Defense have shown interest in elements of a mapping or sequencing project.

Discussions have begun informally on how the agencies might coordinate their efforts, but formal negotiations have yet to begin. Jealousy among the agencies invites speculation that international cooperation is more likely than inter-agency cooperation.

DoE kicked off this year's debate about the human genome with a meeting last March (see *Nature* 321, 371; 1986). Charles DeLisi, director of DoE's Office of Health and Environmental Research, insists it was never his department's intention to sequence the genome, but many came away with that impression after the March meeting. DeLisi says his agency will continue to support activities directed at an ordered set of overlapping cosmids, improved database management and automation of sequencing and cloning technologies. Of the \$10-\$20 million the department plans to spend over the next five years, one quarter will be available to universities and others outside the department.

At the start of the year, says George Palade of Yale University, biologists were divided into two camps on the question of sequencing the human genome; those who said "do it now", and those who said "don't do it". Now, Palade sees a consensus forming, with ideas surfacing on how to achieve what is admittedly a desirable goal.

The issue facing NIH, says Wyngaarden, is "should we devote resources to sequencing the human genome even if it means limiting other research activities?"

Joseph Palca

Doubts about Department of Energy

Washington

ALTHOUGH the Department of Energy's interest in a human genome sequence or mapping project came as a surprise to some, it is not uncharacteristic of the department's work. In addition to its role in GenBank, two Energy Department laboratories have constructed a chromosome-specific DNA library. But some prominent scientists have doubts about the level of biological talent at Department of Energy laboratories.

Their concern arises from contamination problems with the department's chromosome-specific libraries. In order to sort the larger human chromosomes, it was necessary to construct hybrids using Chinese hamster DNA. This inevitably leads to some contamination, because hamster and human DNA cross-react. But some argue that the amount of contamination makes the libraries of questionable value.

Lee Hood of the California Institute of Technology says he is "appalled and de-

pressed" at the efforts of Los Alamos and Livermore National Laboratories to make the libraries. Larry Deaven of Los Alamos National Laboratory, who co-directed the project, does not agree that the work was improperly done. He says the contamination problem is closer to 10-15 per cent, although for chromosome 11 he admits as much as 40-50 per cent hamster DNA is present. But Deaven says he would not make any significant change in procedures. Tom Shows of Roswell Park Memorial Institute and a member of Deaven's scientific advisory panel agrees that the Energy department's effort was as good as existing technology permitted. By taking on a project for wide-scale distribution — more than 1,000 libraries have been handed out — Shows says the effort faced a difficult task in maintaining purity. But Shows' data indicate the contamination of the chromosome 11 library is closer to 60 per cent, a discouragingly high figure.

Joseph Palca

Contraceptive vaccines

Trials in India launched

New Delhi

CLINICAL trials of two locally developed birth control vaccines have started in India. Two hundred sterilized women volunteers recruited at five centres will receive the vaccines in a phase one trial to determine side effects, dose rates and relative efficiencies. The effectiveness of the vaccines in blocking pregnancy in unsterilized women will be assessed in a phase two trial, to begin by the end of the year.

The Indian Council of Medical Research, which has given top priority for immunological research in contraception, is supervising the trials of the vaccines, developed by a team led by Dr G. P. Talwar, director of the National Institute of Immunology in New Delhi. According to Talwar, the vaccines are improved formulations of the subunit hCG (human chorionic gonadotrophin) vaccine he developed twelve years ago, which went into clinical trials in five countries in 1976-78.

The vaccines prevent pregnancy by producing antibodies against hCG, a hormone secreted by the pre-implantation embryo which is essential for establishment and maintenance of pregnancy. Because the hormone is a 'self' protein, the trick is to make the hormone antigenic by tagging it to a suitable carrier. Talwar's early vaccine consisted of the purified and processed β -subunit of hCG coupled to tetanus toxoid (TT) carrier. The β -hCG-TT vaccine was tested in 63 sterilized women in India, Finland, Sweden, Chile and Brazil under a programme supported by the International Committee for Contraceptive Research in New York. It was found to be devoid of side effects, but further trials were abandoned because of poor antigenicity, wide variations in the amount of antibodies produced and low response in one-quarter of test subjects.

Talwar says these limitations have now been overcome, first by 'associating' the

α -subunit of ovine luteinizing hormone (oLH) with hCG to raise its antigenicity and, second, by using two carriers, TT and cholera toxin-B (CHB), to take care of hyporesponders. The idea of using mixed carriers, according to Talwar, is that those who do not respond to one carrier will respond to the other.

What Talwar has now is a cocktail of antigens and carriers. Two vaccines have been standardized by the permutation and combination of these. One of the formulations under trial is a mixture of oLH-hCG-TT and oLH-hCG-CHB. The second formulation consists of a mixture of oLH-TT-CHB and hCG-TT-CHB. Both these formulations have been extensively tested on baboons.

Reporting the findings at a recent international conference on contraception in New Delhi, Talwar said his new vaccines produced twenty times more antibodies than were obtained with his first vaccine. He said the animals responding poorly to the single carrier responded extremely well to mixed carriers. In the absence of booster shots, the animals conceived and delivered normal babies, showing that the effect of the vaccine is reversible. With the launching of trials in India, there is now a race between Talwar and the US-Australian group that, last February, began testing a similar vaccine in Australia under the auspices of the World Health Organisation (see *Nature* 317, 288; 1985). That vaccine, developed by Vernon Stevens of Ohio State University, is based on the carboxy-terminal region of hCG conjugated to diphtheria toxoid.

The results of the Australian trial will be published first, but Talwar is sceptical of its outcome as he believes that carboxy-terminal β -hCG is only a poor immunogen. The carboxy-terminal peptide, however, is specific to hCG and, unlike hCG, does not cross-react with LH. According to Talwar, cross-reaction with LH is beneficial and not harmful. In fact, Talwar's first vaccine has now been cleared for human trials in the United States by the Food and Drug Administration.

Although Talwar was the first to put hCG vaccine into human trials in 1974, he lost the race because of controversies that cropped up after he jumped the gun. In a hurry to beat his competitors, he vaccinated six unsterilized women with hCG-TT vaccine in 1976 when its efficacy was still in doubt. Two of the women became pregnant, World Health Organisation withdrew support and questions of ethics raised by the Indian scientific community forced him to go back to the laboratory and animal trials.

Talwar says he now has more animal data to show his vaccines are efficient and safe. The babies born to the two immunized women are perfectly normal, a point stressed by Talwar to prove the vaccine's safety.

K.S. Jayaraman

Soviet energy

Resurgence of hydropower

SMALL is beautiful once again in Soviet energy planning. Several small hydroelectric stations, mothballed in the mid-1960s, are now being recommissioned. In the post-Chernobyl energy crunch, caused not only by the loss of the capacity of three of the four Chernobyl reactors (No. 1 is now operational again) but also by the temporary withdrawal from service of all RBMK reactors to upgrade their safety systems, every kilowatt of capacity counts. Furthermore, now that the Russian ecologists and historians have won their campaign to halt plans to divert the northern-flowing rivers to the arid lands of central Asia, energy planners are taking a closer look at unused generating capacity.

Small hydroelectric plants were a feature of Soviet energy policy for many years. Some seem to have been built for no better reason than that a region could report that it had no "idle" rivers whatsoever. But others played an effective role in providing electricity for isolated towns of up to 70,000 inhabitants. In some cases, however, the use of lowland rivers with a relatively small gradient led to the flooding of valuable farmland and caused widespread disturbance to the water table of large areas.

By the mid-1960s, the energy planners were firmly committed to the idea of a relatively small number of giant power stations, and, ultimately, a Union-wide grid and some 300 small stations with capacities up to 30,000 kW were phased out. Now, some of these small stations are to be brought back into service and the

first of them, the Kalinin station in Kirgizia, was recommissioned in August. A survey is being made of possible sites where small hydropower stations can be renovated or constructed, particularly in places where there are already reservoirs so that no further flooding or risk to the water table is likely.

Soviet economists are reported to be enthusiastic about the small stations. In remote areas, hydroelectric power can be up to 50 times cheaper than that from oil-fired stations. Small hydroelectric plants, moreover, can be automated, so that there is no need to provide a service infrastructure of housing, schools and shops.

But in the Armenian SSR, for example, where 16 small hydroelectric stations are scheduled for reconstruction during the coming year, hydroelectric power has a bad reputation. An over-enthusiastic hydropower programme in the 1960s led to the depletion of Lake Sevan, the lake which, together with Mount Ararat, forms the symbolic heartland of Armenian tradition. Eventually, in an early victory for Soviet ecologists, the Armenian hydroelectric stations were closed down, the energy plans for the republic were switched to nuclear power and a multi-million-ruble tunnel was drilled through the mountains to replenish Lake Sevan. The announcement on Moscow radio last month, that by recommissioning the small stations the generating capacity of Armenia will have been increased by over 60 MW, will undoubtedly have caused renewed apprehensions.

Vera Rich

European microchips

Commission on horns of dilemma

THE European Commission finds itself in a cleft stick in trying to protect the interests of European semiconductor microchip manufacturers. What was deemed to be a breach of international trading agreements by Japan and the United States may constitute a serious case of dumping. The core of the difficulty is the pact signed in July between Japan and the United States that controls the prices of mass-produced memory chips (see *Nature* 320, 567; 1986).

The Commission's immediate reaction was to challenge the validity of the US-Japan agreement on the grounds that it is tantamount to a cartel and thus a breach of the rules of international trade as outlined by the General Agreement on Tariffs and Trade (GATT). The agreement also has features that seem to threaten the independence or commercial sovereignty of European member states.

The chip agreement was concluded at the eleventh hour, at a time when Japanese manufacturers were under the threat that tariffs would be imposed on memory semiconductors imported from Japan into the United States. The major US manufacturers, backed by the Department of Commerce, had insisted that chips were

being dumped on the US market.

The agreement went further in specifying that the chip trade would be monitored in third countries, including those of Europe, through which Japanese microchips might be imported into the United States. This seemed to the commission to be an extension of the extra-territoriality rules by which the United States has in recent years sought to regulate the use made of US technology in Europe. Under the agreement, even exports to the United States by off-shore subsidiaries of Japanese companies will be controlled, while US chip-makers are guaranteed favourable access to the Japanese domestic market.

The European Commission fears that the agreement will favour US over European companies. It also fears that one effect of the agreement will be to increase

the price of chips in Europe, thus giving a marketing edge to Asian and Pacific Basin manufacturers.

In the event, prices in Europe have not risen with the increases in the past few months in the United States, which is an embarrassment for the commission. Rather, some of the major European manufacturers of chips, through the European Electronic Components Manufacturers' Association (EECMA), have accused Japan of dumping cheap microchips in Europe, and have registered a complaint to that effect with the commission.

According to the secretary general of EECMA, Neville Lyons, "if a country takes on itself the dumping of goods, it should not be surprised if some counteraction is taken". EECMA insists that it does not seek an agreement similar to the US-Japanese accord but the imposition of a trade tariff. The complaint will formally be made to the commission within the next few weeks.

Bill Johnstone

Nobel chemistry

Molecular beams come through

THE Nobel committee for chemistry has singled out for this year's prize the technique for studying chemical reactions by the collision of molecular beams containing molecules whose mechanical state is well defined. The prize is shared by three people responsible for the crucial steps in the development of the technique, Professor Dudley Herschbach from Harvard University, Dr Yuan Tseh Lee, now at the University of California at Berkeley, and

enormously improved the scope of application of the technique by the use of mass spectrometers for the diagnosis of the reaction products. Polanyi, again independently, developed an infrared technique for the diagnosis of reaction products.

In the past 15 years, the applications of reacting molecular beams has been enormously extended, although the complexity of the equipment necessarily restricts the use made of the technique to a com-

paratively small number of laboratories. It now appears possible to prepare beams of neutral molecules with well-determined kinetic energy which are homogenous in the sense that the vast majority of molecules are in well-defined

Soviet Antarctic research

THE Soviet Union's Antarctic geophysical programme will continue and will even be "considerably expanded" in spite of the recent disappearance of the *Druzhnaya-1* station. According to Dr Evgenii Korotkevich, deputy director of the Institute of Arctic and Antarctic Research in Leningrad, there had been plans to work for many more years with the station, which was situated in the Weddell Sea region, near the Filchner and Ronne glaciers. The station was not normally staffed during the southern winter, and its disappearance (either down a crack in the ice or by drifting out to sea) was observed by satellite at the beginning of October. *Druzhnaya-1*, situated about 3 km from the edge of the Filchner glacier, came to grief when a slice of the glacier's 100-km-wide ice foot broke up and began moving out to sea.

Since it was established in 1975, the station had been the main base for Soviet geological and geophysical research, including the investigation of the Weddell Sea area to ascertain if oil or gas-bearing deposits were present. The 170 geologists due to arrive at the station as part of the 32nd Soviet Antarctic expedition at the end of October will now be deployed to the *Druzhnaya-2* and Soyuz stations. Further stations in the Antarctic will also be opened in the near future.

Vera Rich



Prizewinners Herschbach (left), Lee and Polanyi.

Professor John Polanyi from the University of Toronto.

Herschbach appears to have been, in 1962, the first to turn earlier tentative experiments with reacting molecular beams into the basis of the now widely practised technique. The first experiments involved the interaction of alkali atomic beams with those of simple molecules such as halogen hydrides. Polanyi's independent contribution, at this early stage, was the development of techniques diagnostic of the state of the reaction products.

For nearly a decade, the application of the technique was restricted to elementary variations of the reactions studied by Herschbach, but early in the 1970s, Herschbach in collaboration with Lee, then a post-doctoral fellow in his laboratory,

states of vibration and rotation. Techniques of beam preparation have improved so that almost any material that will vaporize without molecular change can be reacted with other materials.

The benefits of the techniques stem largely from the lack of ambiguity in the interpretation of the results which are obtained. One physical chemist explained this week that the feasibility of being able to account for the activation energy of a chemical reaction not by some kind of 'Boltzmann smearing' over all possible states but by the identification of those states that would react and those which were indifferent helped to fix ideas in people's minds. In due course, he said, it would be necessary to rethink the way students are taught about reaction mechanisms. □

Nobel physics

Electron microscopy acclaimed

ERNST Ruska, who receives half of this year's Nobel prize in physics, was the first to build an electron microscope, the design of which was essentially that of current machines. He has worked on electron microscopes ever since. His most important work began when he was a graduate student and was completed by 1938, by which time he was working for Siemens and Halske.

As a graduate student, Ruska worked at the Technical University of Berlin, where with Knoll in 1931 he showed that it was possible not only to form but also to magnify an electron image. The relative contributions of Ruska and Knoll to this work have never been clear, but Ruska pursued the development of the electron microscope while Knoll instead became interested in television.

Ruska, for a time, was alone in his belief that a useful electron microscope could be built, but he teamed up with Bodo Von Borries and began to construct a prototype machine. Von Borries was more of a businessman and seems to have been responsible both for the key patent relating to the iron pole piece lens in 1933 and for attempts to win commercial backing.

In the end, the development was financed by Siemens and Halske, which gained the right to manufacture the machine, the first of which was made in 1938. Ruska and Von Borries worked in a converted bakery throughout the Second World War. Their laboratory escaped bomb damage but was looted by the Soviet army. Ruska continued working at Siemens until the 1950s, when he joined the Fritz Haber Institute, part of the Max Planck Institute in Berlin, at which he has remained.

It is unclear why it has taken so long for Ruska, who will be 81 on Christmas Day, to be awarded a Nobel prize. Although there seems no doubt that he was the brains behind the first electron microscope, the claims of other participants, including Knoll and Von Borries, must at first have complicated matters, as would have the circumstance that the first patent belonged to Rheinhold Rüdenberg, apparently with little justification.

For many years, the Nobel committee may have balked either at awarding a prize to Ruska alone or at deciding who else should share it. Could it really be that the committee has only now found a solution to its problem in the form of an invention that allowed two quite unrelated people to take the other half of the prize?

If the award to Ruska is overdue, this cannot be said of the other half of this year's prize, which goes to Gerd Binnig and Heinrich Rohrer of the IBM Research Laboratory in Zurich, Switzerland, for the

invention of the scanning tunnelling microscope. The first instrument was built in Zurich between 1979 and 1981, and the paper which, according to one colleague, "won them the prize", was published in *Physics Review Letters* only in 1983.

The scanning tunnelling microscope, the first scanning instrument to resolve single atoms, uses the phenomenon of vacuum tunnelling. A voltage applied between a very sharp needle (the tip) and the surface under study allows electrons to



Nobel laureates Ruska (left), Binnig and Rohrer.

tunnel from sample to tip; because the tunnelling effect requires the electron wavefunctions in tip and sample to overlap, the tunnelling current is a very sensitive function of the tip-sample separation. This allows variations of surface structure to be mapped within a vertical resolution of 0.07 Å and a lateral resolution of 1–3 Å. The microscope can also act as a probe of the electronic structure of a surface.

Although scanning tunnelling microscopy is still in its infancy, the technique is being used by more than 50 university and industrial research groups. So far, it has been applied most widely in the field for which it was invented, the study of semiconductor surfaces, but new applications are rapidly emerging. The microscopes have now been used in air at atmospheric pressure, and have even been shown to work under water.

The only severe limitation to the technique is that it cannot be used to study insulating materials. Binnig, who is setting up a new laboratory at IBM Munich, has started to tackle this problem with an instrument he calls the atomic force microscope. Here the tip is positioned so as to maintain a constant force (instead of tunnelling current) between tip and sample; the rapid decrease of interatomic force with distance gives

this instrument a vertical resolution of less than 1 Å and a lateral resolution of 30 Å. The atomic force microscope can be used to study conductors or insulators.

Rohrer, aged 53, and Binnig, 39, have each been awarded an IBM fellowship, a renewable appointment, given to honour "sustained technical achievement", which allows the recipient to pursue his own project for up to five years.

Peter Newmark & Laura Garwin

European research

Tangled framework programme

FRANCE is having trouble in keeping up even present levels of support for research organized by the European Commission in Brussels; Britain and West Germany would support "modest" increases; and other countries would support the doubling requested by the Commission. This was the spectrum of national positions shortly before the meeting of the European Council of Research Ministers on Tuesday. The meeting was the last before the self-imposed 9 December deadline by which Britain, President of Council, plans a decision on the Commission's £6,200-million, five-year proposal for a 'framework programme' for research, and while the gap might appear unbridgeable, the solution is in the Commission's hands, according to British officials.

The real block, in the British view, is that the Commission refuses to disentangle its 'framework' for research into individual packages that could be identified as particular programmes (such as ESPRIT, the European stimulation programme for research on information technology). The

Commission, led by its president Jacques Delors, insists on a unanimous decision on the aggregated 'framework' package, where individual nations cannot easily identify and support specific programmes. Only when the framework is agreed, he argues, will the Commission show its hand, by which time the European Single Act, the replacement for the Treaty of Rome which was the foundation of the European Economic Community, should be in place, allowing decisions on individual programmes by majority voting in the council of ministers. Britain is arguing, with the support of most other countries, that this is unrealistic. The Commission should disaggregate the framework programme, allowing real bargaining to begin among the various programmes and national interests. After that, a decision on financing could be reached relatively quickly, British officials believe.

Early this week, however, it seemed likely that the Commission would stick to its guns, with an impasse in December increasingly probable. Robert Walgate

Investigating the paranormal

SIR—David Marks' Commentary (*Nature* 320, 119; 1986) is a comprehensive presentation of the arguments that believers in the universal validity of scientific materialism could raise against investigation of the paranormal. But some of the arguments are circular in the sense that they are valid only if scientific materialism is taken as the only possible base.

Marks quotes the opinions of individual philosophers as definitive and accepted truth without indicating that others hold different opinions on the same questions which are more sympathetic to the paranormal cause and which cannot be dismissed as inferior.

For example, Marks quotes Flew's opinion¹ about the impossibility of an after-life without referring to the philosophical difficulties of that opinion (LeShan²) and without mentioning the existence of empirical³⁻⁵ claims to the contrary. It may be true that the investigation of near-death experience (an example of paranormal investigations) can be interpreted on a materialistic basis just as in Marks' account of parapsychology, but to take Flew's argument as a solid conclusion is a serious misinterpretation of the philosophical standing of his argument.

Similarly, Marks quotes Scriven's⁶ philosophical opinion about the impossibility of non-materialistic evidence, using this as a definitive and a settled argument against paranormal investigation. But other philosophers (for example, Polanyi⁷ and Feyerabend⁸) have different opinions about both the feasibility and desirability of such *ad hoc* modifications of materialism.

Marks makes the surprising claim that there are no theories of paranormal investigation. Does he simply mean that there are no materialistic theories of the paranormal? There are, of course, philosophically cautious theories of the paranormal⁹ and different models of objective transcendence^{10,11}. Or does Marks mean that paranormal investigation has as yet no paradigm (in the Kuhnian sense) that would serve as an example to guide further paranormal investigation? This is true, but only because the investigation of the paranormal is not yet a mature science.

Such confusions arise because Marks does not distinguish between philosophical and scientific theories. There are many who will also reject his assumption that scientific thinking is necessarily analytical thinking¹² and that scientific analytical thought will inevitably replace other perceptions of the world. It is an open question whether such development is likely to be a permanent trend^{13,14}, let alone whether it is desirable.

Because so many of Marks' assumptions are based on problematical philo-

sophical outlooks, his final conclusion is too severe. The interpretation of paranormal investigation is a promising endeavour in the pre-paradigm stage. The question of whether it is a scientific endeavour depends on our perception of what is science: was the investigation of electricity before Franklin properly called science?

MICHAEL WAGNER
Searle Research & Development,
4901 Searle Parkway,
Sokokie, Illinois 60077, USA

1. Flew, A. in *Science, Pseudoscience and Society*. (eds Hanen et al.) 55–75 (Wilfred Laurier University Press, 1980).
2. LeShan, L. *How to Meditate* 14th edn 19–24; 69–72 (Bantam, New York, 1984).
3. Moody, R. *Life after Life* (Bantam, New York, 1976).
4. Sabhom, M. *Recollections of Death* (Harper & Row, New York, 1982).
5. Ring, K. *Life at Death* (Coward, McCann & Geoghegan, New York, 1982).
6. Scriven, M. in *Philosophy and Psychical Research*. (ed. Thakur, S.C.) 181–194 (Allen & Unwin, London, 1976).
7. Polanyi, M. & Prosch, H. *Meaning* (University of Chicago Press, 1972).
8. Feyerabend, P. *Against Method* (New Left Books, London, 1975).
9. LeShan, L. *The Medium, the Mystic and the Physicist: Toward a General Theory of the Paranormal* (Viking, New York, 1974).
10. Smith, H. *The Forgotten Truth* 38–48; 63–73 (Harper & Row, New York, 1976).
11. Bentov, I. *Stalking the Wild Pendulum* 3rd edn 95–142 (Bantam, New York, 1971).
12. Wagner, M. *Nature System* 6, 53–76 (1984).
13. Broch, H. *The Sleepwalkers* (Random House, 1948).
14. Toynbee, A. *A Study of History* (Oxford University Press, 1946).

Systematics as science

SIR—J.A. Barnett (*Nature* 322, 599; 1986) exemplifies many users of the results of systematics who do not realize that systematics is a science in itself, not a mere service to others. If I worked on yeasts rather than on mammals, perhaps I could invert the usual request and ask Barnett, or someone else, for a small favour.

"My work on genus X has encountered difficulties that your techniques can probably ameliorate. I therefore request that you stop whatever you may be doing and sequence several genes from each of the 23 species for which I enclose samples with this request. This should, of course, be at your own expense. It does not matter if your own work is on something else and you have no interest in these species or genes; your knowledge of the techniques is what is important for my work. If you are especially cooperative I will consider mentioning you in the acknowledgements of my paper."

This is actually not an outrageous caricature of some requests to systematists. Systematists are seen as technicians because, indeed, identification and nomenclature are at that level and that is where interactions of systematists with others usually occur.

It is nevertheless true, and not "romantic confusion", that systematics deals with the underlying structure of evolution. As

with any living science, systematics makes progress. A by-product of this progress is that classifications and names, the results most visible to others but nevertheless only superficial manifestations of the real body of knowledge, can change. The changes affect other people. But so do advances in any science. They should be welcomed as advances of a dynamic and living science.

LEIGH M. VAN VALEN
Biology Department (Whitman),
University of Chicago,
915 E. 57 Street,
Chicago, Illinois 60637, USA

Wetlands research

SIR—In his review of our book, *Sweet Track to Glastonbury* (*Nature* 323, 120; 1986), G.J. Wainwright commented on the need for a Wetland Resource Centre to represent the broad range of scientific interests that wetland archaeology encompasses.

To meet this need, the University of Exeter has established a Wetland Archaeology Research Project to co-ordinate relevant work, to provide advice and to act as a central focus for the exchange of information about archaeological survey, excavation, analysis and conservation. Many archaeologists, environmentalists and conservators have already become associates of the project, and it is hoped to expand the list to include all those interested in the potential and the achievement of wetland archaeology, including those concerned with wetland conservation in its widest sense.

J.M. COLES
B. COLES
Wetland Archaeology Research Project,
Department of History and Archaeology,
University of Exeter,
Exeter EX4 4QH, UK

Trieste confusion

SIR—I would like to point out an error in John Maddox's article "New ways with interatomic forces" (*Nature* 322, 769; 1986). The reference to "... a group based at the University of Trieste (F. Ercolessi, E. Tosatti, M. Parrinello)" is wrong: these three scientists all belong to the International School of Advanced Studies in Trieste.

The school was created in 1978 and is independent of the University of Trieste; it is an "experimental" institution in the Italian university system inspired partly by the Scuola Normale Superiore of Pisa, the École Normale Supérieure of Paris and the Princeton Institute for Advanced Studies.

PAOLO BUDINICH
International School for
Advanced Studies,
Strada Costiera 11,
34014 Trieste, Italy

Turbulence assails fifth force

Claims that gravitational measurements in the 1920s can embody evidence for novel short-range forces could have a more humdrum explanation.

THE saga of the 'fifth force' continues, but in an uncertain direction. The story, less than a year old, goes back to the re-analysis by Fischbach *et al.* (*Phys. Rev. Lett.* **53**, 3; 1986) of the measurements by Eötvös and associates in the 1920s of the gravitational attraction of several different kinds of material by the Earth. Although the old data have been taken for the past half century as a proof that gravity depends only on mass and not, for example, on chemical composition, they have errors (of the order of one part in 10,000) which appear to vary systematically with chemical composition. Is that correlation real, or just an artefact?

The little bombshell devised by Fischbach *et al.* began with the observation that the systematic errors in the old data appear to vary with the baryon content, in this context the nucleon content, of the materials used by Eötvös. The point is that the baryon content, or in this context simply the nucleon content, varies with atomic number; the greater the atomic number, the greater the number of nucleons per unit mass of material, largely on account of the nuclear packing fraction but also because the greater ratio of neutrons to protons in heavier nuclei means that the small electron mass makes a lesser contribution to the total.

But why should baryon number be what matters? In the contemporary fashion, Fischbach *et al.* interpreted the systematic variation of apparent gravitational attraction as the consequence of the exchange between baryons of low-mass, perhaps massless, particles; the shadowy creature known as the axion is one candidate. The outcome of the re-interpretation was that there would be a force resembling gravitation between all pairs of objects that would nevertheless fall off exponentially with distance (over a range measured in tens of metres) and that would be determined not by the product of the two masses but by the product of their baryon numbers, which is nearly, but not quite, the same thing (see *Nature* **319**, 173; 1986).

Inevitably, the appearance of Fischbach *et al.* began a hearty buzz. Reactions have ranged from disbelief to solemn speculation about the implications of the new force, the fifth force (supplementing the two nuclear forces, electromagnetism and gravity), for theories of particle physics.

Among experimentalists, there has been a spate of proposals for making more sensitive measurements than were pos-

sible in 1922. Some of the best proposals would not be directly concerned with variations of chemical composition, but are designed to separate the short-range fifth force (which is, by the way, repulsive) from long-range gravity by measurement of the variation of apparent gravitational forces with distance in the neighbourhood of known density inhomogeneities, as on the face of a cliff.

Suggestions that apparent gravity may not be exactly that specified by Newton's inverse-square law are by no means new. More than a decade ago, for example, Daniel R. Long described a series of measurements apparently showing that gravitational forces are an increasing function of distance over a range of tens of centimetres (*Nature* **260**, 417; 1976), but several attempts to confirm this conclusion have not done so. More recently, F. D. Stacey from the University of Queensland has argued that measurements with a gravimeter in an Australian borehole provide evidence for departures from newtonian attraction at distances of the order of metres (see Stacey, F. D. & Tuck, G. J., *Nature* **292**, 230; 1981), although the promised repetition of that measurement in a tract of deep ocean (where density variations are probably better known) seems not yet to have materialized. At the beginning of this year, it may be thought, the time was ripe for radical innovation in the field.

Now the small and esoteric community of those concerned with the measurement of gravity will be in a tizzy all over again. S. Y. Chu from the University of California, Riverside, and R.H. Dicke from Princeton University have now put out the claim that the systematic errors in the original Eötvös experiments are not the consequences of a fifth force but of a much more mundane phenomenon: thermal convection (*Phys. Rev. Lett.* **57**, 1823; 1986). What they have seized on is the simple fact that the physical bulk of a sample suspended on an Eötvös balance will depend on the atomic number of its constituents, as will the physical cross-section of the sample presented to thermally driven movements of the atmosphere in the neighbourhood of the equipment, simple draughts in the laboratory.

With only a little special pleading, Chu and Dicke are able to show that the conventional forces thus generated might account for the systematic errors. Fischbach *et al.* admit that there is a case to

answer. It is interesting that Robert Dicke who has the reputation of an iconoclast, should on this occasion be defending the *status quo*.

Experimentalists whose grant applications are even now being judicially considered need not despair. Few will risk taking sides on the evidence available. The difference between the two groups demonstrates most clearly of all that the original data-set, however convincing it may have been in the 1920s as a proof that the gravitational attraction between masses is independent of their chemical composition, is entirely inadequate for the purposes for which it is now being used, the attempt to understand the origins of the systematic errors.

The Eötvös experiments involved a torsion balance designed so that dissimilar materials could be suspended from the two arms. There appear to have been three series of measurements, one in which the gravitational attraction of the Earth for three materials (Cu, Mg-Al alloy and "snakewood" (*sic*)) was compared with that for platinum, another in which an unreacted mixture of AgSO₄ and FeSO₄ was compared with itself after chemical reaction and a third series in which the attraction of pairs of materials are compared directly with each other.

In total, there are only nine pairs of measurements, two of which were omitted from the original analysis by Fischbach *et al.* while a third (involving radium) is omitted from the analysis by Chu and Dicke. Chu and Dicke have done well to extract from the data a correlation between the surface area of the original samples and the departures from the inverse-square law recorded in the 1920s. Fischbach *et al.* rightly say that there are features of the analysis by Chu and Dicke that cannot be unambiguously resolved at this late stage; the general opinion is that there will have to be more experiments.

To judge from what people say, there will be no shortage of them. Fischbach *et al.* have put gravity measurements on the map again. Such a development can only be worthwhile; it may even contribute to the easing of what remains a blot on the face of physics, that the gravitational constant ('big' *G*) remains uncertain to about one part in 5,000. **John Maddox**

A fuller account of this issue, by Alvaro de Rújula of CERN, will appear shortly in the News and Views pages.

Theoretical chemistry

Back to the valence bond

from Roy McWeeney

THE nineteenth-century theory of chemical combination — the valence bond — may be making a comeback. Nowadays students are taught that the molecular orbital theory is the proper means to understand chemical bonding, and that the valence bond theory is more or less defunct. But on page 699 of this issue, David L. Cooper, Joseph Gerratt and Mario Raimondi claim that valence bond theory, when properly implemented, leads to a picture of the delocalized double bonds in benzene — the textbook example

hold them together. In the linear combination of atomic orbitals (LCOA) approximation, ϕ is approximated as a mixture ($\phi = c_1\chi_1 + c_2\chi_2$) of the atomic orbitals and the electron density is then a weighted sum of the three terms indicated in Fig. 2. The interpretation of the bond (though not the calculation itself) is 'classical' because the small magnetic interactions of spins are negligible at the thermochemical (non-spectroscopic) level.

It is not immediately obvious how a valence bond approximation can lead to a

in the determination of the electronic energy. But the idea of linking energies with alternative spin couplings took root in the 1930s and has been used (and mis-used) ever since.

More refined wavefunctions for hydrogen (for example, valence bond functions with ionic terms in which one atomic orbital holds 2 electrons and the other none; or molecular orbital functions with one or both electrons in the anti-bonding molecular orbital; Fig. 1b) give computed electron densities in perfect agreement at the 'basis set limit' — the best that can be done using only two 1S atomic orbitals. Only by greatly extending the basis is it possible to approach 'exact' solutions of the Schrödinger equation.

The valence bond versus molecular orbital argument is therefore not about which theory is 'right' but rather about which is easier to use, which gives most rapid convergence, and which gives the nicest interpretation of the bonding.

Cooper and collaborators have now considered the prototype of all aromatic molecules, benzene, re-opening old disputes about the description of the double bonds. Here the six carbon atoms provide six π electrons, occupying atomic orbitals with one lobe above and one below the molecular plane. In ethylene (C_2H_4), two such π electrons lead to the bond, which reinforces the C—C single bond and yields the double bond $C=C$, but in benzene there are only enough electrons (three pairs) to provide three π bonds; the question is where to put them.

Kekulé's proposal (in 1865) that the double bonds might somehow oscillate between two alternative structures (Fig. 3a) is embodied in a valence bond wavefunction with two main components, each with a spin-pairing scheme corresponding to such a structure; for benzene this is often called a resonance hybrid of the two Kekulé structures. The description is radically different from that used in molecular orbital theory, where pairs of electrons are put in the molecular orbitals shown in Fig. 3b and delocalization of the bonds arises because the molecular orbitals (and hence the electron density) are smeared over all six C—C links. Both theories agree that benzene possesses six identical 'partial' double bonds; but to bring their results into numerical agreement no less than 175 valence bond structures (2 Kekulé, 3 Dewar and 170 'ionic', or a like number of molecular orbital

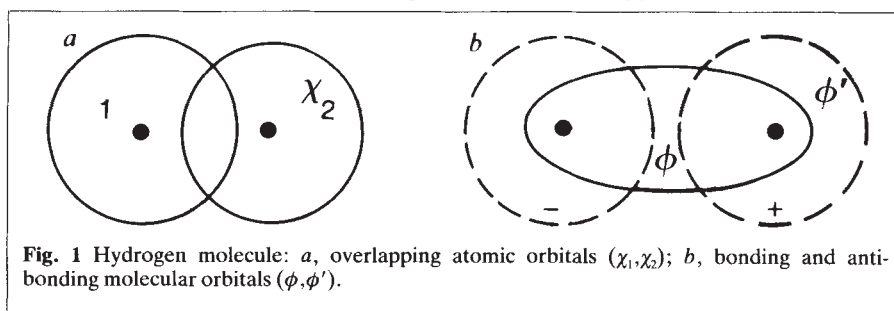


Fig. 1 Hydrogen molecule: a, overlapping atomic orbitals (χ_1, χ_2); b, bonding and anti-bonding molecular orbitals (ϕ, ϕ').

— which is unbeatable in elegance, economy and precision.

What most students are not told is that when both theories are carried rigorously to their logical conclusion their results must be mathematically identical: both seek an approximation to the many-electron wavefunction and both use orbitals (that is, wavefunctions for single electrons) as 'building blocks', and it is only in the simplest approximations of both types that disagreement is possible. Whereas the valence bond approach uses the atomic orbitals of the bonded atoms, molecular orbital theory assigns the electrons to molecular orbitals which extend over the whole molecule. Orbitals of both kinds are shown schematically in Fig. 1 for the hydrogen molecule.

The simplest molecular orbital approximation assigns the two electrons to the bonding (low-energy) molecular orbital, say ϕ , with spins opposed to give a singlet state; and the resultant electron density is $P = 2\phi^2$, which is large between the nuclei and provides the attractive forces which

similar picture, for with electron '1' in χ_1 , and electron '2' in χ_2 , it is not possible to get any overlap density (Fig. 2). But the electrons may change places (each taking the position, and spin, previously belonging to the other); and when the wavefunction is allowed to describe a 'mixture' of both possibilities, the calculated electron density does contain an overlap term — although with slightly less weight than in the molecular orbital approximation. In the singlet ground state, the spins of the two electrons are opposed (antisymmetric coupling) but the orbital factor in the wavefunction is symmetric (unchanged if the electrons change places), and it is this symmetry which leads to an overlap density with a positive weight factor. The sign is reversed on adopting an antisymmetric orbital factor, and the bonding is turned into antibonding, but the spins must then be parallel-coupled (symmetric case) to give overall antisymmetry of the wavefunction — which then describes an excited triplet state.

The spin coupling thus only reflects the symmetry of the orbital part of the wavefunction; it is a 'flag' and has no direct role

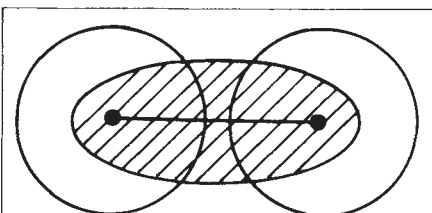


Fig. 2 Hydrogen molecule: orbital (χ_1^2 and χ_2^2) and overlap ($\chi_1\chi_2$) terms in the electron density.

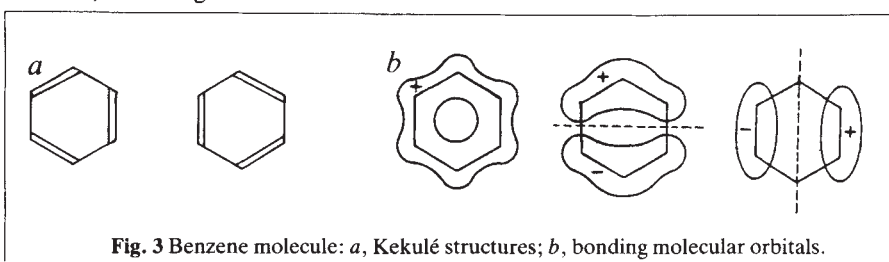


Fig. 3 Benzene molecule: a, Kekulé structures; b, bonding molecular orbitals.

terms, must be used.

What Cooper and co-workers have shown is that the valence bond picture is improved if the atomic orbitals are 'optimized' (each one spread out slightly towards its neighbours). Then the two Kekulé structures alone give very nearly the same result. In other words, as they are progressively refined, the two ways of constructing electronic wavefunctions (one using orbitals localized around individual atoms, the other using orbitals spread over the whole molecule) approach the same end point at very different rates; a few Kekulé structures based on well-chosen localized orbitals may apparently be equivalent to a much longer expansion of the electronic wavefunction in terms of molecular orbitals.

The new approach thus does not in practice lead to a new electronic structure

for benzene, for the term 'structure' really means the electron density, and that can be obtained by either method with equal precision. Neither, unfortunately, will the method provide an easy route to similar *ab initio* calculations on large molecules, an arduous task whatever method is used. What it does do is to provide a remarkable vindication of many of the qualitative ideas formulated long before the era of the computer, expounded at length in the books by Pauling and by Wheland; ideas which have long been a cornerstone in chemistry and which will continue to stimulate the imagination of both experimental and theoretical chemists. □

Roy McWeeney is Professor of Theoretical Chemistry in the Department of Chemistry and Industrial Chemistry, University of Pisa, Via Risorgimento 35, 56100 Pisa, Italy.

Nitrogen fixation

Finding a time and a place

from A.E. Walsby

NITROGEN fixation presents a problem for the cyanobacteria because the enzyme that catalyses the reduction of dinitrogen is inactivated by oxygen and oxygen is generated by photosynthesis. It is essential, therefore, that these two processes are separated from one another. In many filamentous cyanobacteria they are separated spatially, photosynthesis into vegetative cells and nitrogen fixation into heterocysts (Minchin, F. *Nature News and Views* 320, 483; 1986). A. Mitsui and his collaborators now describe on page 720 of this issue how in a unicellular cyanobacterium the two processes are separated temporally into different parts of the cell cycle.

The idea that heterocysts provide anaerobic conditions for nitrogen fixation in filamentous cyanobacteria was first put forward (Fay, P. *et al. Nature* 220, 810; 1968) at a time when all the strains for which there was unequivocal proof of nitrogen fixation were known to possess heterocysts. But within a year the first report of nitrogen fixation by non-heterocystous cyanobacteria appeared (Wyatt, J.T. & Silvey, J.K.G. *Science* 165, 908; 1969), clearly establishing nitrogenase activity under aerobic conditions, in a species of *Gloeotheca* — then called *Gloeocapsa* — a unicellular form distinguished by its concentric layers of sheaths (see figure). If this unicellular organism could fix nitrogen under aerobic conditions, as R. Rippka *et al. (Arch. Mikrobiol.* 87, 93; 1971) confirmed, why should other strains of cyanobacteria require heterocysts?

The first insight into how *Gloeotheca* managed to separate nitrogen fixation

from photosynthesis was provided by P.M. Millineaux, J.R. Gallon and A.E. Chaplin (*FEMS Microbiol. Lett.* 10, 245; 1981) who found that in cultures of this organism grown under a cycle of alternating 12 hours light and 12 hours dark, 95 per cent of the nitrogenase activity occurs in the dark. There is thus a temporal separation of these two incompatible processes; during the dark period respiratory activity reduces the oxygen tension and helps protect the nitrogenase, whereas in the light period oxygen is produced by photosynthesis and the nitrogenase is inactivated. New nitrogenase must be synthesized at the start of the next dark period.



Phase-contrast light micrograph of *Gloeotheca* sp. showing concentric sheaths laid down at successive cell divisions. (Bar; 10 μ m.)

This explanation would hold for cells growing in the natural (or artificial) light-dark cycle but how would it account for nitrogenase activity under continuous illumination? The answer to this has now been provided by Mitsui and his collaborators.

These authors, working with another unicellular nitrogen-fixing cyanobacterium, a species of *Synechococcus*, performed experiments very similar to those of Millineaux *et al.* but managed to obtain a culture of synchronously dividing cells. As with *Gloeotheca* there was little or no nitrogenase activity in the 12-hour light period and activity reached its peak near the middle of the dark period. This peak followed soon after the sharply synchronized cell division that occurred after 4 hours of darkness.

When, after three such light-dark cycles, the culture was maintained in continuous light a synchronized cell division occurred 24 hours after the previous division and was also followed by a rise in nitrogenase activity. These results clearly establish that the nitrogenase activity is associated primarily with a particular phase in the cell cycle rather than with light intensity alone. Mitsui and collaborators have also shown that within this cell cycle there is also a cycle of photosynthetic capacity, reaching a maximum during the light phase and a minimum of nearly zero in the dark phase at a time coinciding with the peak in nitrogenase activity.

Another aspect of the process still remains to be explained: how does the cell protect its nitrogenase from oxygen diffusing into the cell from the outside? Such a problem is common to other non-photosynthetic bacteria that fix nitrogen under aerobic conditions (see Bothe, H. in *The Biology of the Cyanobacteria* eds Carr, N.G. & Whitton, B.A. p.87; Blackwell, Oxford, 1982). In the case of *Gloeotheca* it may be significant that sister cells, which retain a degree of synchronous development, are kept in proximity by the concentric arrangement of sheaths (see figure). They may in this way provide for themselves a synchronized micro-environment that maintains a lower oxygen tension during the nitrogen-fixing phase of the cell cycle.

We now, however, at last have a crucial part of the explanation of how such unicellular cyanobacteria are able to fix nitrogen in continuous light: each cell independently continues through its cell cycle, switching first to nitrogenase activity in the absence of oxygenic photosynthesis, and then to photosynthesis in the absence of nitrogen fixation. The temporal separation of the processes thus continues with endogenous timing by the cell division cycle. □

A.E. Walsby is Professor of Botany at the University of Bristol, Bristol B58 1UG, UK.

Materials science

Melting and the surface

from Robert W. Cahn

ALTHOUGH freezing is a complicated process, there is general agreement about its essential features; but melting is still a mystery, and opinions about its origin abound. Recent experiments, however, seem to narrow the options and it is hard to avoid the conclusion that normal melting is initiated by a continuous vibrational lattice instability at the solid surface or at a solid-surface interface.

Ubbelohde¹, in a survey of models of melting, stresses the Lindemann criterion, formulated in 1910, according to which melting is a vibrational instability released when the root-mean-square amplitude of vibration reaches a critical fraction of the interatomic distance. In a comprehensive recent survey, Boyer² concludes that a lattice shear instability is the precipitating feature for melting, whereas Cotterill³ emphasizes catastrophic defect (dislocation) generation as the crucial factor. (I have discussed⁴ the role of vacancies in melting in a previous *News and Views* article.) Couchman and Jesser⁵ made an early attempt at a synthesis of these apparently contradictory models: these authors pointed out that "the dislocation theory of melting, in essence the idea that [T_m is] the temperature at which the free energy of formation of dislocations is identically zero, implies, as does the Lindemann hypothesis, the vanishing of the static shear modulus". What is most important about their paper, however, is the recognition that melting is a surface-initiated process.

One clear indication of the role of the surface is that until recently it was believed to be impossible to superheat a crystal above its melting point except by arranging to heat a solid from the interior⁶, so that the interior reached the melting temperature before the surface did; even then, the attainable superheating was always much smaller than the supercooling that can be imposed metastably on a melt.

This idea has now been disproved by an experiment⁷ in which small (~0.15-mm diameter) silver monocrystal spheres were coated with gold, as $T_m(\text{Au})$ is greater than $T_m(\text{Ag})$, and superheated for periods of the order of 1 min by up to 24 K above $T_m(\text{Ag})$; the limited time results from diffusion of gold from the coating into the silver. This experiment is different from that of Williamson *et al.*⁸ who used picosecond electron diffraction in association with pulsed laser heating to show that thin, solid aluminium films can be superheated by more than 1,000 K but only for periods of ≤ 1 ns; here there is only tran-

sient but not metastable superheating.

Another body of evidence that assigns a crucial role in melting to the solid surface relates to the change of melting temperature with particle size. Buffat and Borel⁹, among others, showed that small particles of Au, Pb and other metals suffer a melting-point depression of up to as much as 30 per cent for particles of 20–30 Å diameter. Hoshino and Shimamura¹⁰ have interpreted this in the light of the well-known fact that the mean square amplitude of vibration at a solid surface is usually 1.5–2 times larger than in the bulk (for example, ref. 11); their theory is based on the idea that the Lindemann criterion for a vibrational catastrophe applies to the surface vibrational amplitudes rather than to the bulk values. Similar ideas have been suggested simultaneously by Pietronero and Tosatti¹².

Another kind of small particle is discussed by Rossouw and Donnelly¹³, who injected argon ions into aluminium at room temperature. Under these circumstances, the Ar forms minute bubbles (mean diameter 2.6 nm) in the form of a crystalline face-centred cubic solid in epitaxy with the Al lattice. The melting temperature of Ar at atmospheric pressure is 84 K. In fact, the diffraction pattern of solid Ar remains detectable up to a temperature of 730 K, that is, a metastable superheating of 480 K!

Precision measurements of diffracted intensities allow mean square vibration amplitudes to be estimated for both Al and Ar, and the striking fact emerges that the Ar amplitudes are smaller than for free solid Ar. Specifically, the Debye temperature, a measure of cohesion deduced from the different intensities, is 139 K for the Ar bubbles compared with 70 K for Ar at atmospheric pressure and 110 K computed from 'free' Ar pressurized in the bubbles. (A high Debye temperature implies a higher cohesion, smaller vibration amplitudes and, generally, a higher melting temperature.) Because the bubbles are so small that more than half the atoms can be regarded as being effectively 'surface' atoms, it follows that the amplitude of surface (more precisely, interface) vibrations is depressed when Ar is in contact with Al (which has a higher Debye temperature of 398 K). Superheating is possible because the surface (interface) Ar atoms are constrained to smaller vibrations than the interior atoms. No doubt a similar explanation can be advanced for the substantial superheating observed for 2-nm hydrogen bubbles in an α -silicon matrix¹⁴.

In a study similar to the argon-in-aluminium investigation, Evans and Mazey¹⁵ injected krypton into titanium at 300 K and again found that the Kr was present in the bubbles as crystals epitaxial with the titanium, this time with a hexagonal close-packed structure. The Kr melts progressively during heating, the smaller bubbles (with smaller Kr lattice parameters and accordingly higher pressures) melting at higher temperatures. The melting of individual bubbles can be detected by high-temperature transmission electron microscopy, and survival of the smallest bubbles to higher temperatures is thus confirmed. The maximum superheating here is much smaller than in the case of Ar: the highest T_m observed is 380 K compared with a pressure-adjusted 'free' T_m of 335 K (the T_m for free Kr at atmospheric pressure is 117 K).

In commenting on earlier work by Evans and Mazey⁶ on Kr bubbles in Cu and Ni, Rossouw and Donnelly remark that the Kr lattice parameter increases (that is, the pressure decreases) after cycling through the melting transition and back; they presumed that this is caused by an inflow of vacancies from the host metal to the molten Kr bubbles. Such an inflow would modify the interface structure. They add that no corresponding change in the lattice parameter of their Ar bubbles in an Al host is observed after a similar melting cycle. This suggests that these bubbles are immune to vacancy inflow and the character of the interface is thus unaltered on cycling through the melting transition. Rossouw and Donnelly's experiment seems to be an unambiguous demonstration of the role of the surface (interface) in controlling the melting process. (In the study⁷ of the melting of Au-coated Ag microspheres, the Ag was not under hydrostatic pressure, and therefore the modest superheating in that experiment is not connected with pressure.) The Ar-in-Al and Au-on-Ag experiments suggest that the observed

1. Ubbelohde, A.R. *The Molten State of Matter: Melting and Crystal Structure* (Wiley, Chichester, 1978).
2. Boyer, L.L. *Phase Transitions* 5, 1 (1985).
3. Cotterill, R.M.J. *J. Cryst. Growth* 48, 582 (1980).
4. Cahn, R.W. *Nature News and Views* 273, 491 (1978).
5. Couchman, P.R. & Jesser, W.A. *Phil. Mag.* 35, 787 (1977).
6. Cormia, R.L., Mackenzie, J.D. & Turnbull, D. *J. Phys. Chem.* 65, 2239 (1963).
7. Däges, J., Gleiter, H. & Perepezko, J.H. *Mat. Res. Soc. Symp. Proc.* 57 (1986).
8. Williamson, S., Mourou, G. & Li, J.C.M. *Phys. Rev. Lett.* 52, 2364 (1984).
9. Buffat, Ph. & Borel, J.-P. *Phys. Rev.* A13, 2287 (1976).
10. Hoshino, K. & Shimamura, K. *Phil. Mag.* A40, 137 (1979).
11. Goodman, R.M. & Somorjai, R.M. *J. chem. Phys.* 52, 6325 (1970).
12. Pietronero, L. & Tosatti, L. *Solid State Comm.* 32, 255 (1979).
13. Rossouw, C.J. & Donnelly, S.E. *Phys. Rev. Lett.* 55, 2960 (1985).
14. Boyce, J.B. & Stutzmann, M. *Phys. Rev. Lett.* 54, 562 (1985).
15. Evans, J.H. & Mazey, D.J., *J. Nucl. Mat.* 138, 176 (1986).
16. Evans, J.H. & Mazey, D.J. *J. Phys.* F15, L1 (1985).
17. Frenken, J.W.M. & van der Veen, J.F. *Phys. Rev. Lett.* 54, 134 (1985).
18. Lipowsky, R. & Speth, W. *Phys. Rev.* B28, 3983 (1983).

superheating depends on epitaxy between the contained solid and the coating/host metal, and it is possible that that an incoherent interface might not permit substantial superheating.

Further evidence concerning the surface role in melting is provided by ion-channelling studies¹⁷ using proton sensing beams on an atomically clean (110) surface of a heated lead crystal. The solid-liquid transition begins 40 K below the equilibrium bulk T_m , and the thickness of the molten surface layer increases dramatically close to the bulk T_m , being as much as 20 layers within a fraction of 1 K of that temperature. This implies second-order (continuous) melting at the surface, and thus raises in general terms the question of the order of transformation of the melting process at a surface; Frenken and van der Veen¹⁷ quote recent theoretical work¹⁸ on atomic ordering at a surface (for example,

of Cu₂Au) which proves that an order-disorder transformation that is first-order in the bulk can be second-order (continuous) at the surface. That prediction has been confirmed by several experimental studies of Cu₂Au. Indeed, Frenken and van der Veen claim, but do not prove, that the Lipowsky/Speth approach¹⁸, when applied to melting, predicts a liquid-film thickness diverging proportionally to $[T/(T_m - T)]$, where T_0 is constant, which is consistent with their own experimental results for lead.

Thus, taking together all the new evidence the almost unavoidable conclusion is that melting does indeed begin as a continuous vibrational instability at the solid surface of a solid-solid boundary. □

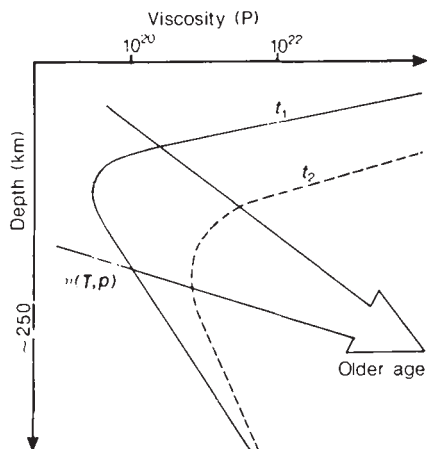
Robert W. Cahn is in the Department of Metallurgy and Materials Science, Cambridge University, Cambridge CB2 3QZ, UK.

Mantle geophysics

Variable viscosity makes waves

from David A. Yuen

BOUNDARY-layer instabilities develop in thermal convection, when the driving buoyant forces are sufficiently strong. In the laboratory one can monitor these hydrodynamic events in constant-viscosity fluids, or in temperature-dependent viscous fluids such as corn syrup or silicone oil. The flow law of mantle materials,



Schematic indication of variation with depth of temperature- and pressure-dependent viscosity $\eta(T, p)$. The times are given by t_1 and t_2 with t_2 greater than t_1 . Arrow denotes the trend caused by the aging of the lithosphere. A low viscosity zone would be present at a depth of around 100 km.

however, depends strongly on both the temperature and the pressure. Up to now, the importance of temperature- and pressure-dependent viscosity has not been widely appreciated, mainly because of the proliferation of laboratory and numerical experiments on constant or temperature-dependent viscosity fluids. Recent work

by W.F. Haxby and J.K. Weissel¹ may have put temperature- and pressure-dependent viscosity back in the limelight.

In their analysis of Seasat altimeter data Haxby and Weissel find lineated patterns of gravity anomalies and sea-surface height variations occurring over the younger portions of the fast-moving plates in the Pacific and Indian oceans. Haxby and Weissel suggest that these lineations, with wavelengths in the range of a couple of hundred kilometres, are caused by small-scale convective instabilities in the upper mantle beneath the oceanic lithosphere at plate ages as young as 5 million years. Previously, convective instability, based on constant viscosity models, has been advocated to explain the flattening of the sea floor from the conductive cooling predictions at plate ages greater than about 80 million years². For secondary convection to be triggered beneath the young sea floor, the effective viscosity locally can be estimated to be about 10^{19} poise, which is substantially less than the mean upper mantle viscosity inferred from glacial isostasy. From the dominant wavelengths of the undulations over young oceanic crust, Haxby and Weissel then reckon from linear stability arguments that the thickness of this convecting layer is about 100 km. Such localized cells may evolve within a low viscosity channel (see figure), which is a consequence of the intrinsic temperature- and pressure-dependent nature of mantle viscosity in the presence of temperature fields near the ridge.

The idea that a layer of pronounced low viscosity is needed to induce at an early age instabilities of the cold upper bound-

ary layer was advanced in 1981³ and confirmed later by numerical computations^{4,5}. Because the thermal boundary layer thickens with age of the sea floor, the low-viscosity zone becomes less pronounced with time (see figure) and moves deeper into the mantle. This downward migration of the low-viscosity well would influence the wavelengths of the gravity anomalies with age. Finite-amplitude calculations⁶ predict the wavelengths would increase more rapidly than the observations. This suggests that these signatures were frozen in the growing elastic portion of the oceanic lithosphere.

Temperature- and pressure-dependent viscosity may also be influential in thinning the lithosphere, caused by the impingement of a hot plume coming from the upward migration of the low-viscosity zone (reverse t_2 and t_1 in the figure). This process would involve a positive-feedback mechanism, as the viscosity well becomes more pronounced and the local viscosity drops down sharply with greater amounts of penetration by the plume⁷. For this reason, the formation of seafloor swells by mantle plumes should be investigated numerically from the point of temperature- and pressure-dependent viscosity, rather than from numerical or laboratory studies⁸ in which a low, constant viscosity fluid is assumed.

Besides its role in upper mantle dynamics, temperature- and pressure-dependent viscosity can be important in several situations involving deep mantle circulation. For example, the adiabatic temperature gradient in the lower mantle is linked strongly to the parameters of the flow law, such as the activation volume and the Grüneisen parameter. Too large an activation volume (greater than 6 ml per mole), would lead to super-adiabatic temperature gradients in the lower mantle⁹.

For geodynamicists to be further convinced of the importance of temperature- and pressure-dependent viscosity in mantle convection, corroboration of these initial exciting findings of small-scale convection at young ages² is needed. More work should show that these gravity perturbations do not have a lithospheric origin. □

1. Haxby, W.F. & Weissel, J.K. *J. geophys. Res.* **91**, 3507 (1986).
2. Parsons, B.E. & McKenzie, D.P. *J. geophys. Res.* **83**, 4485 (1978).
3. Yuen, D.A., Peltier, W.R. & Schubert, G., *Geophys. J. R. astr. Soc.* **65**, 171 (1981).
4. Yuen, D.A. & Fleitout, L. *Phys. Earth planet. Inter.* **34**, 173 (1984).
5. Buck, W.R. thesis, MIT (1984).
6. Buck, W.R. & Parmentier, E.M. *J. geophys. Res.* **91**, 1961 (1986).
7. Yuen, D.A. & Fleitout, L. *Nature* **313**, 125 (1985).
8. Olson, P.M. & Nam, I.S. *J. geophys. Res.* **91**, 7181 (1986).
9. Quaren, F., Yuen, D.A. & Saari, M.R. *Geophys. Res. Lett.* **13**, 38 (1986).

David A. Yuen is at the Minnesota Supercomputer Institute and Department of Geology and Geophysics, University of Minnesota, Minneapolis, Minnesota 55455, USA.

Biophysics

Use of deuterated proteins in nuclear magnetic resonance

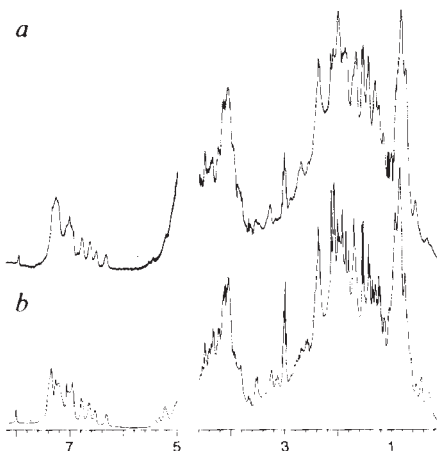
from Paul Rösch

PROTON nuclear magnetic resonance (NMR) spectroscopy provides information on the tertiary structures of small compounds of biochemical interest such as peptides, small proteins or pieces of DNA. It is also well suited to monitor qualitatively structural changes induced by complex formation of different molecules such as enzymes and their substrate ligands. Attempts to measure the change of ligand conformation on binding to a protein is severely impaired because in most cases proton resonances from the protein tend to obscure all or many of the ligand resonances, especially if the ligand molecule is itself a peptide or a protein. In addition, the magnetic dipolar interactions between ligand and protein protons prevent either conformation from being determined independently, rendering structure determinations of such systems virtually impossible. A recent study by Seeholzer *et al.*¹ attempts to distinguish structurally between the inactive and active forms of an enzyme. The authors studied the complex of the eukaryotic calcium-dependent regulatory protein calmodulin and the bee-venom peptide melittin. The calmodulin/melittin complex is probably not physiologically significant, but has potential as a model of the calmodulin/target-protein interaction.

To observe the proton NMR spectrum of the bound peptide Seeholzer *et al.*¹ used perdeuterated calmodulin and protonated melittin as complex constituents. This allows the spectrum of the ligand to be observed without interference from the protein resonances. The use of partially or completely deuterated compounds in ¹H-NMR to observe protonated ligands is by no means a new approach, first publications on the subject appearing as early as 1968 (refs 2,3).

What is novel about the work of Seeholzer *et al.* is that they used a completely deuterated protein of animal origin. They obtained it in the straightforward way of cloning the bovine calmodulin gene into *Escherichia coli* and subsequently growing the bacteria in a perdeuterated medium. Because only the signals from the protonated 26-amino-acid peptide melittin appear in the spectrum of its complex with the 148-amino-acid protein calmodulin, Seeholzer *et al.*¹ could also deduce the spectrum of the protein in the complex by use of difference spectroscopy between the deuterio-calmodulin/melittin complex and the fully protonated calmodulin/melittin complex.

Studies of structural changes induced by complex formation between proteins and their target or substrate molecules are among the most interesting in protein biophysics. In most cases, information on protein-bound molecules is obtained in two ways: either by indirect observation of the influence of complex formation via detection of excess ligand molecule, making use of the fact that the spectrum of the uncomplexed ligand in the presence of the complex-forming molecule is disturbed by the influence of intermediate complex formation; or by monitoring nuclei present



Comparison of ¹H NMR spectra of free calmodulin (a) and the calmodulin/melittin complex (b). (From ref. 1.)

exclusively or predominantly in the ligand, such as ³¹P in the substrates of phosphoryl-transferring enzymes⁴ or ¹³C, ¹⁵N and ¹⁹F of isotopically labelled or enriched ligand molecules. In fact, one of the co-authors of the publication by Seeholzer *et al.*¹, Mildred Cohn, devoted much of her career to observations of ³¹P NMR spectra of such enzyme-bound substrates, which provided valuable information on protein-ligand interactions.

To gain insight into the structure of protein-bound molecules it is in general necessary to analyse their proton NMR spectra. The observation of such spectra of ligands in the complex can be accomplished as described above. In a way, this procedure is rather similar to the substitution of perdeuterated solvents in NMR (and other branches of spectroscopy) to enhance the observability of sample resonances close to the solvent peaks. As a bonus, one also gets rid of the magnetic dipolar interactions between protein protons and ligand protons, so that only intrapeptide interactions need consideration in

the liquid-NMR context. This is very useful for the final structure determination by two-dimensional NMR⁵ because the ligand can be studied independently.

From the overall appearance of the spectrum of bound melittin¹ the presence of structured strands in the peptide can be inferred. Melittin is in an extended form as a monomer in D₂O, but assumes more structure in deuterio-methanol^{1,6} and when bound to deuterated lipids⁷. As the authors themselves state, a detailed analysis of the structure of calmodulin-bound melittin will only be possible using two-dimensional NMR.

Judging from the quality of the spectra, obtaining two-dimensional spectra in D₂O should be straightforward. It may be even easier than collecting high-quality two-dimensional nuclear Overhauser-effect spectra of a peptide of this size free in solution (see the recent discussion in *News and Views*⁸), because the motional correlation times are longer in the complex. In contrast, obtaining two-dimensional spectra of the ligand involving exchangeable protons in H₂O, which is necessary for the sequential assignment of resonances, poses a problem, because the protein and ligand resonances are expected to overlap in this region. Many resonance assignments have already been made for the free peptide⁶.

Seeholzer *et al.* present a simple and straightforward application of genetic engineering methods to enhance the effectiveness of NMR. Cloning of calmodulin into *E. coli* was the necessary prerequisite for obtaining the perdeuterated species. Unfortunately, it is not possible to grow every particular strain of bacteria in perdeuterated medium. Neither is it very easy to incorporate specifically all types of amino acids in protonated or deuterated form into bacteria. On the other hand, genetic engineering methods now allow sequence-specific substitution of amino acids at the gene level (site-directed mutagenesis). Spectral assignments of single amino-acid residues and studies of the structural influence of single amino-acid substitutions by NMR should thus be possible. This technique of 'site-directed NMR' holds great promise. □

1. Seeholzer, S.H., Cohn, M., Putkey, J.A., Means, A.R. & Crespi, H.L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3634 (1986).
2. Markley, J.L., Putter, I. & Jardetzky, O. *Science* **161**, 1249 (1968).
3. Rosenberg, R.M., Crespi, H.L. & Katz, J.J. *Biochim. biophys. Acta* **175**, 31 (1969).
4. Rösch, P. *Prog. Nucl. Magn. Reson. Spectrosc.* **18**, 123 (1986).
5. Bax, A. & Lerner, L. *Science* **232**, 960 (1986).
6. Lauterwein, J., Brown, L.R. & Wüthrich, K. *Biochem. biophys. Acta* **622**, 219 (1980).
7. Lauterwein, J., Bösch, C., Brown, L.R. & Wüthrich, K. *Biochim. biophys. Acta* **556**, 244 (1979).
8. Kabsch, W. & Rösch, P. *Nature News and Views* **321**, 469 (1986).

Paul Rösch is in the Department of Biophysics, Max-Planck Institute for Medical Research, 6900 Heidelberg 1, FRG.

Developmental biology

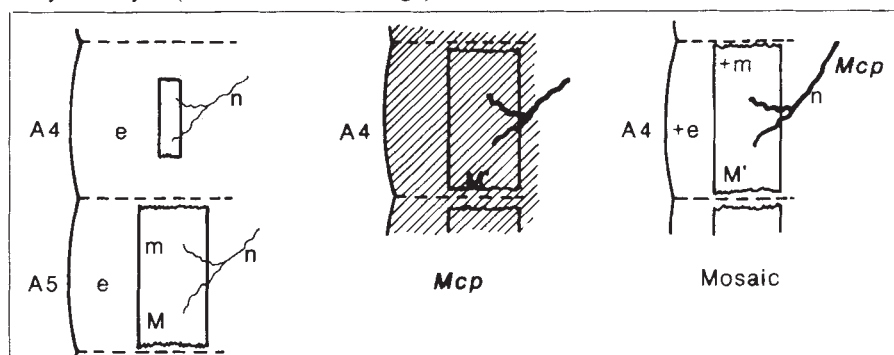
Nervous excitement over insect muscle patterns

from Vernon French

PATTERN formation in insect mesoderm has been ignored until recently, being wrongly assumed either to occur by processes identical to those patterning the overlying epidermis, or to be a passive reflection of them. Recent studies show that muscle patterns are built by a fascinating array of interactions within the mesoderm, with the epidermis and, surprisingly, with the nervous system. They are unlikely to be ignored again.

There is considerable evidence that in the fruitfly *Drosophila* groups of cells become segmentally determined in the early embryo (late blastoderm stage),

accordance with their particular combination of active selector genes so that, in a chimaeric animal, cells mutant for a selector gene will make structures of an inappropriate compartment in an otherwise normal animal. Exactly which structures are made by particular cells then depends on interactions with neighbouring epidermal cells, but there are no indications of major influences from other tissues. Similar studies of segmental determination within the mesoderm give rather different results and suggest that muscle patterns depend on influences from other tissues.



Dorsal muscle patterns in *Drosophila* abdominal segments (summary of results in ref. 8). Wild-type (+) males have a large dorsal muscle (M) only in segment A5, whereas *Mcp* mutant males have this large muscle (M') also on their transformed segment A4. Nuclear transplantation gives mosaic animals in which muscle M' can form on A4 from wild-type mesoderm (m) attached to wild-type epidermis (e), provided that the innervation (n) is *Mcp* mutant.

although segmental organization only becomes visible later with the appearance of grooves in the epidermis, a repeating pattern of neuroblasts and condensations in the underlying mesoderm¹. If a single ectoderm cell is genetically marked at blastoderm stage, the resulting clone does not cross between segments or between anterior and posterior compartments within a segment. Furthermore, ectoderm cells transplanted into another body segment at or after late blastoderm continue to form structures of their segment of origin. The analysis of the phenotypes of homoeotic mutants and the *in vivo* localization of the corresponding gene transcripts strongly indicates that segmental determination is accomplished by the precisely patterned activity of various regulatory 'selector' genes² such as those of the bithorax complex³ and the Antennapedia complex⁴. In fact, the initial unit of determination may well not be the segment but the parasegment, comprising the posterior compartment of one segment and the anterior one of the next⁵. Epidermal cells develop autonomously in

The mesoderm is formed by the early invagination of a strip of ventral blastoderm cells and, as a first step in investigating its segmental determination, Lawrence⁵ tested for lineage restrictions by initiating clones at or after blastoderm stage. Clones were mutant at the enzyme locus *sdh* and were distinguishable from wild-type cells after appropriate staining. Clones in the adult thorax were precisely restricted to separate muscle sets corresponding to the three segments (but, unlike epidermal clones, not further confined to an anterior or posterior compartment). Lineage restriction alone does not prove segmental determination, however, as groups of cells may become clonally distinct just because they become spatially separated. In normal development the adult head and thorax form from several separate imaginal disks so, for instance, the wing-disk epidermis forms part of the dorsal thorax and its myoblasts form the associated muscles.

Lawrence and Brower⁶ implanted a disk fragment into the abdomen and found that the myoblasts dispersed and could con-

tribute to any nearby host muscles. This result is compatible with the traditional view of insect mesoderm, derived largely from the results of early surgical experiments (see ref. 5), that the muscle pattern depends only on instructions from the overlying ectoderm. In the disk implants, wandering myoblasts could now be instructed by abdominal epidermis and would develop accordingly. But these experiments are not conclusive, as muscles are formed by cell fusion and the nuclei of implant-derived myoblasts could be reprogrammed by the cytoplasm of the developing host muscle. Hence muscle pattern could, after all, depend on the segmental determination of the mesoderm cells and thus on their selector gene activity⁷.

Selector genes

The role of selector gene activity in muscle patterning has recently been explored by Lawrence and Johnston⁸ in an elegant study of the large dorsal longitudinal muscle which is normally formed only in the fifth abdominal segment (A5) of male *Drosophila* (see figure). In the bithorax complex mutant *Mcp*, segment A4 is transformed to A5 and, in males, forms the large muscle. The mutation could directly affect the mesoderm cells of A4 or influence the muscle pattern via an effect on some other cells, and this question was resolved by studying muscle patterns in a wide range of mosaic animals. Nuclei were injected at early precellular stages from *Mcp* mutant donors into host embryos which were mutant for *sdh* but had a wild-type bithorax complex. The injected nuclei populated a variable small region of the blastoderm cell layer and eventually formed a patch of detectable *Mcp* cells in the adult fly.

The results show clearly that the ectopic muscle is not formed because of altered instructions from the overlying epidermis. There were several male mosaics where the large muscle formed in segment A4 beneath wild-type epidermis or the normal small A4 muscle formed under *Mcp* epidermis, so muscle pattern is not correlated with the genotype of the epidermis. Surprisingly, the genotype of the mesoderm cells themselves does not determine the muscle pattern in segment A4, as donor-derived *Mcp* cells could form the small A4 muscles while host-derived cells could form the large A5-type muscle. Many of these mosaics were gynandromorphs where donor and host were of different sexes and, again, the large 'male' muscles could be formed in segments A4 or A5 in cases where either the overlying epidermis or the muscle itself is composed entirely of female cells.

The mosaics provide intriguing evidence that it is the sex and segmental determination of the innervating motor neurones that controls which type of muscle is

formed. The genotype of the motor neurone endings could be scored only in some animals, but gave an impressive correlation with muscle pattern. Hence the large muscle forms in segment A5 if the innervation is from male neurones, but not if innervation is female, and the ectopic large muscle forms in segment A4 if the motor neurones are male and *Mcp*, but not if they are either female or male but wild type for the bithorax complex. There are many instances where the survival or complete differentiation of insect muscles depends on innervation⁹, but this is the first demonstration that the form of a muscle (its size and its points of attachment) may be controlled by its innervation.

Mosaic analysis

The analysis of mosaics demonstrates that it is not the genotype of the myoblasts or of the dorsal epidermis to which they attach which determines the form of the muscle, but that "it is almost impossible to prove positively what is responsible"⁸. It may not be the innervating motor neurones themselves which influence muscle development but rather some other cells originating from the same area of the blastoderm and therefore reliably having the same genotype in these mosaics. Hence the critical interaction could be with cells elsewhere in the ventral nerve cord (or possibly in the ventral epidermis, which also forms from the embryonic neurectoderm).

Analysis of the mosaics is also unable to reveal when the interaction occurs, and this problem is compounded by our considerable ignorance of the details of the formation of these muscles. The origin of embryonic muscles has been carefully described in the locust where each muscle starts as one or a few large mesoderm cells called muscle pioneers¹⁰. The pioneers attach to the epidermis at the eventual sites of muscle insertion and, as the embryo grows, numerous small mesoderm cells fuse on and build the syncytial muscle fibres¹¹. Muscle pioneers are contacted at an early stage by growth cones of pioneer neurones and one now wonders whether this contact is important for the future development of the muscle.

Drosophila pupae are very different from locust embryos. The adult abdominal epidermis develops from small groups of cells (the histoblast nests) which divide rapidly in the pupa, replacing the degenerating polyploid larval epidermis. It seems that the adult muscles are similarly formed from small numbers of myoblasts which divide and fuse in groups, but it is not known whether the muscles form *de novo* (and, if so, whether there are specific pioneer myoblasts) or are assembled on the remains of larval muscles. The interaction between neurones and myoblasts of segments A4 and A5 could therefore

Fritz Lipmann (1899–1986)

FRITZ LIPMANN died on 24 July, shortly after learning that his latest research grant would be approved: he and others in his laboratory were studying the mechanism of regulation of cell growth and transformation. Recently, he had returned to his early interest in the energetics of phosphoproteins and he showed that the phosphorylation of a tyrosine residue by ATP, catalysed by an oncogene-encoded phosphokinase, is readily reversible, thereby demonstrating the energy-rich nature of a phosphorylated tyrosine residue.

Lipmann is largely responsible for our present understanding of the energetics of living systems at the molecular level, as described in his classic 1941 review "Metabolic Generation and Utilization of Phosphate Bond Energy" published in volume 1 of *Advances in Enzymology*. The importance of his contribution is so fundamental that it has become hard to recognize — the 'energy-rich' bond is understood by every student as the driving force for virtually all biological processes, in one sense or another. This was not always so. Lipmann's descriptions of 'high-energy' compounds or group activation were strongly attacked by several physical chemists and it took some years before this semantic uproar died away and the significance of these compounds became widely accepted.

Lipmann has probably identified more kinds of energy-rich compounds than anyone else. In 1939 he reported that acetyl phosphate could transfer phosphate to ADP and later he showed that it could transfer an acetyl group to various acceptors in the presence of bacterial enzymes. His interest in active acetate led to an examination of the acetylation of sulphanilamide and to the identification of the acyl group carrier coenzyme A (CoA), leading

to his Nobel prize in 1953.

An examination of the ATP-dependent synthesis of acetyl-CoA provided the first demonstration of ATP cleavage at the α - β bond to expel pyrophosphate in activation reactions. This demonstration was rapidly followed by the identification and synthesis of active carbamate, carbamyl phosphate, and the active sulphates adenosine-5'-phosphosulphate and 3'-phosphoadenosine-5'-phosphosulphate. Lipmann made many important conceptual and experimental contributions to the mechanisms of protein and peptide synthesis, including the characterization of the energy-rich bond of 2'/3'-aminoacyl transfer RNAs and stimulatory protein factors. He elucidated a remarkable mechanism for the synthesis of polypeptide antibiotics in the absence of nucleic acids through a sequence of specific protein-bound thiol esters of amino acids and peptides.

Fritz Lipmann was born in Königsberg. He received a medical degree in Berlin in 1924 and a doctorate in chemistry in 1927. He worked with Otto Meyerhoff and Albert Fischer and spent several years in Copenhagen and at the Rockefeller Institute before his appointment to the Department of Surgery at Massachusetts General Hospital in 1941. His best-known work was carried out in the research laboratories of this institution, where he was appointed Professor of Biochemistry at Harvard Medical School in 1949. He returned to Rockefeller University in 1957 and became Professor Emeritus in 1970. In 1966 he was awarded the National Medal of Science by President Johnson. He characterized his career as "following one's instinct without knowing exactly where it will lead". His instinct was famous for knowing where to go. William P. Jencks

occur at early embryonic stages, influencing pioneer cells for the adult muscles, or producing covert differences in larval muscles which will later act as a template. Alternatively, there could be a later interaction between the developing adult muscles and their innervating motor neurones.

Muscle pattern

The nuclear transplantation experiments strongly indicate that the differences between the dorsal muscles of male segments A4 and A5 are controlled by segmental determination within the nervous system⁸ but, of course, it does not follow that all features of insect muscle pattern arise in this way. Muscles interact with the epidermis, attaching to particular sites (even if these have been moved or duplicated¹²) and this specific recognition must be vital in generating complex patterns of interweaving muscles, as in the

*Drosophila thorax*³. It is also likely that the muscle pattern depends on some level of segmental determination within the mesoderm itself. The selector genes, whose patterns of activity control ectodermal segmental determination, are also expressed stably within the somatic mesoderm, although in a simpler pattern. Hence the *Ubx* region of the bithorax complex is expressed in a complex pattern within the epidermis and nerve cord of parasegments 5–13, but in a uniform way over mesodermal parasegments 6–12 (which form the abdominal muscles)³.

Of course the sustained patterns of selector gene activity may have no direct function in the mesoderm but this seems improbable, especially as a different selector gene, *engrailed*, seems to have no role in the mesoderm and is barely expressed there¹³. Furthermore, some puzzling features of muscle transformation in bithorax complex mutants can be largely

explained if selector genes do function in the mesoderm. The almost perfect transformation of meta- to mesothoracic epidermis, the 'four-winged fly', occurs in the multiple mutant *abx, bx, pbx*, and this genotype also similarly transforms the ventral nervous system^{14,15}. Critically, the adult¹⁶ and larval¹⁷ muscles of the metathorax are not transformed (as would be expected if their pattern derived from the overlying epidermis or the innervation), but the larval muscles of segments A1 and A2 have recently been shown to be transformed to the metathoracic pattern¹⁷. Because the *Ubx* region (within which these mutations map) is expressed in the ectoderm of the metathorax but not in its mesoderm³ (derived from parasegment 5), the ectodermal and muscle transformations are likely to result from separate direct effects of impaired selector gene function in the two tissues¹⁷.

Of course, more devious explanations of the muscle transformations can be suggested. The inducing element of innervation could arise from a more anterior segment, or partial homoeotic transformations of the ectoderm could alter only its own differentiation in the metathorax and only its inducing abilities in segments A1 and A2. Mosaic experiments similar to those of Lawrence and Johnston⁸ would be required to eliminate these possibilities but, for now, the correspondence between sites of gene expression and areas of transformation argues strongly for a direct mesodermal role for selector genes, leading to at least an abdomen/thorax distinction within muscle pioneer cells. Further segmental differences in muscle patterns could arise through effects of other selector genes (for example, *Antp*, active in mesoderm only in parasegment 5; see ref. 4), through differences in the epidermal pattern of attachment sites or, as shown for segments A4 and A5, through differences within the nervous system⁸. □

1. Lawrence, P.S. & Martinez-Arias, A. *Nature* **313**, 639 (1985).
2. Garcia-Bellido, A. *Ciba Fdn Symp.* **29**, 161 (1975).
3. Akam, M.E. & Martinez-Arias, A. *EMBO J.* **4**, 1689 (1985).
4. Martinez-Arias, A. *EMBO J.* **5**, 135 (1986).
5. Lawrence, P.A. *Cell* **29**, 493 (1982).
6. Lawrence, P.A. & Brower, D.L. *Nature* **295**, 55 (1982).
7. Lawrence, P.A. & Johnston, P. *Cell* **36**, 775 (1984).
8. Lawrence, P.A. & Johnston, P. *Cell* **45**, 505 (1986).
9. Nüesch, H. *Comprehensive Insect Physiology and Biochemistry* (eds Kerkut, G.A. & Gilbert, L.I.) **2**, 425 (1985).
10. Ho, R.K., Ball E.E. & Goodman, C.S. *Nature* **301**, 66 (1983).
11. Ball, E.E., Ho, R.K. & Goodman, C.S. *Devl Biol.* **111**, 383 (1985).
12. Williams, G.J.A. & Caveny, S. *J. Embryol. exp. Morph.* **58**, 13 (1980).
13. Ingham, P., Martinez-Arias, A., Lawrence, P.A. & Howard, K. *Nature* **317**, 634 (1985).
14. Green, S.H. *Nature* **292**, 152 (1981).
15. Teugels, E. & Ghysen, A. *Nature* **314**, 558 (1985).
16. Lawrence, P.A. *Trends neurosci.* **8**, 267 (1985).
17. Hooper, J.E. *EMBO J.* **5**, 2321 (1986).

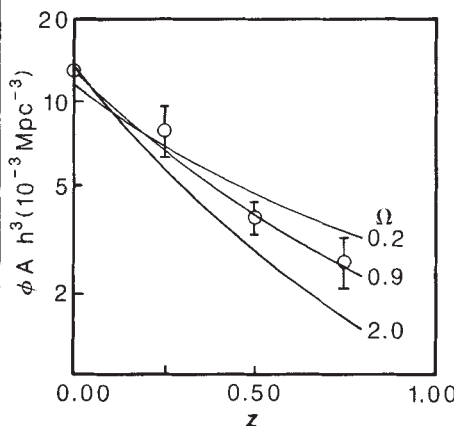
Vernon French is in the Department of Zoology, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK.

Cosmology

Is omega equal to unity?

from Joseph Silk

COSMOLOGISTS are obsessed with the idea that Ω , the density parameter of the Universe, defined to be the ratio of the mean density to the critical closure value, is precisely equal to unity. Were it not so, the argument goes, the initial conditions of the big bang must have been highly contrived, to within 1 part in 10^{60} , to allow the Universe to expand to its present dimension. Moreover, inflationary cosmology provides a physical reason for why Ω should have been so precisely tuned to the value of unity: all curvature, the manifestation of deviations of Ω from unity, was exponentially inflated away during the first 10^{-35} seconds of the cosmic expansion.



Plot of the parameter which measures the deviation in the expected number density of galaxies resulting from the curvature of the Universe versus redshift (z). Points at $z = 0.25, 0.5$ and 0.75 , Loh and Spillar's derived centres of galaxies; solid lines, predictions.

A new measurement of Ω over a large region of the Universe has just been performed by Edwin D. Loh and Earl J. Spillar (*Astrophys. J. Lett.* **307**, L1; 1986). Their result is close to 1.

Astronomers have traditionally measured a value for Ω that is much closer to 0.1 than to 1. This has not deterred cosmologists, whose theoretical studies have mostly focused on the implications of a universe with $\Omega = 1$ for large-scale structure and for the nature and detectability of the dark matter. The measurements of Ω that have hitherto been performed sample a fairly local region of space, out to 10 or 20 megaparsecs. (For comparison the nearest large galaxy, Andromeda, is 0.67 megaparsecs away.) The standard cosmological tests for the curvature of space do sample almost the entire observable Universe, but have hitherto been imprecise in determining Ω because of the uncertain effects of galactic evolution. It is entirely possible that a component of dark matter, invisible to astronomers because it is non-

luminous and uniformly distributed over this relatively small length scale, could dominate the mass density of the Universe, thereby allowing Ω to approach unity when the Universe is fairly sampled over sufficiently large scales.

The Universe has now been probed out to a redshift of about 0.5, or to a scale of order 1,000 megaparsecs by Loh and Spillar, who develop a novel technique for measuring redshifts of a substantial number of galaxies, based on an approach pioneered by R. Kron and D. Kos. The method consists of taking charge-coupled device images of a deep field in six narrow-band filters, ranging in wavelength from 4,250 to 9,000 Å. This yields six colours which provide what is claimed to be a unique photometric measure of the galaxy redshift. The spectral energy distribution of a galaxy varies with wavelength, there being in particular a pronounced spectral break caused by metal absorption near 4,000 Å in the rest frame. At greater distance from us, the redshift increases and the spectroscopic signature marches through the filter sequence. If distant galaxies have similar intrinsic energy distributions to nearby galaxies, this then provides a measure of redshift that is accurate enough to do a cosmological test.

The test consists of measuring the number density of galaxies as a function of redshift (see figure). The number density is equal to its euclidean value if $\Omega = 1$ and space is not expanding. The effect of the expansion reduces the number counted in a static space by a factor that measures the contraction of the proper volume element at earlier epochs. Curvature acts to reduce this further, if the Universe is positively curved, and to increase the relative volume element if the Universe is negatively curved. A low- Ω universe corresponds to a cosmological model in which the curvature of space is negative, and one expects the number counts of objects to be enhanced relative to an $\Omega = 1$ model by about 50 per cent at a redshift of 0.75 if $\Omega = 0.2$.

By estimating redshifts for 1,000 field galaxies with a median redshift of 0.5, Loh and Spillar argue that they have measured the galaxy number density to better than about 20 per cent. Fitting to the luminosity function of galaxies brighter than a specified flux limit enables them to classify the same objects as galaxies at both low and high redshift. Calibration with spectroscopic redshift determinations of galaxies in clusters at $z = 0.4$ suggests that spectral evolution, which could bias their redshift determinations, is unimportant. If they

are not systematically losing (or gaining) galaxies at high redshift because of photometric errors or uncertainties, then Loh and Spillar have succeeded in measuring the volume element as a function of redshift. It does not matter whether the mass in the Universe is luminous or dark. This measurement of the geometry of the Universe yields a determination of $\Omega = 0.9^{+0.7}_{-0.5}$ (95 per cent confidence limits) and is consistent with the inflationary prediction of an Einstein-de Sitter cosmology ($\Omega = 1$).

This result is especially noteworthy because it is the first, seemingly unambiguous astronomical determination of Ω that yields a value of unity on large ($\sim 1,000$ megaparsecs) scales. Before accepting this as the new cosmological gospel, however, notes of caution have been urged. Redshift errors may be systematically underestimated because the 4,000-Å spectral indicator has only been calibrated against cluster galaxies, whereas this fea-

ture is known to be weaker for the field galaxies that dominate the Loh-Spillar sample (Richard Ellis, University of Durham, personal communication). There may be systematic errors in photometry that make the smaller, more distant galaxies appear brighter relative to the nearby galaxies, thereby mimicking the effect of high Ω (S.D. Jorgovski, personal communication). Dust in distant galaxies may also affect the photometric redshift determinations. These effects need to be critically evaluated, presumably by time-consuming spectroscopic studies, before the $\Omega = 1$ result can be considered to be definitive. For now, however, it remains a tantalizing possibility: perhaps observational cosmologists (and other astronomers) should pay more heed to theorists. □

Joseph Silk is in the Astronomy Department at the University of California, Berkeley, California 94720, USA.

Mathematics

The Waring experience

from Ian Stewart

In 1770 Joseph-Louis Lagrange proved a long-sought result in number theory, the four-squares theorem: every whole number is a sum of four perfect squares. Earlier that year Edward Waring, Lucasian Professor of Mathematics at Cambridge, had published his *Meditationes Algebraicae*. It contained a far-reaching generalization, stated without proof: every whole number is a sum of 9 cubes, 19 fourth powers "and so on". The implied conjecture, that for all n the number of perfect n th powers needed to represent any whole number is finite, became known as Waring's problem. The finiteness was proved by David Hilbert in 1909 using an identity in 25-fold multiple integrals, but his method did not yield the actual number of n th powers required. Other methods have been developed to estimate their number, and various workers have improved the upper bounds until they have become quite respectable. R.C. Vaughan has now obtained substantial improvements to these bounds when n lies between 5 and 9 (*Proc. Lond. Math. Soc.* **52**, 445; 1986 and *J. Lond. Math. Soc.* **33**, 227; 1986).

More precisely, let $g(n)$ be the smallest number such that every whole number is a sum of $g(n)$ n th powers. Then Lagrange's result shows that $g(2) \leq 4$. Because the number 7 actually requires all four squares ($7 = 1^2 + 1^2 + 1^2 + 2^2$ is the best we can do), $g(2)$ is exactly 4. Waring's conjecture is that $g(3) = 9$, $g(4) = 19$ and that $g(n)$ exists and is finite for all values of n .

He almost certainly hit on these numbers by trial and error. A search of all

numbers up to, say, a few thousand will show that at most 9 cubes appear to be needed; and only 2 numbers, 23 and 239, require all 9. In fact only 15 numbers seem to need 8, the largest of these being 454. In *An Introduction to the Theory of Numbers* (4th edn; Oxford University Press, 1960), G.H. Hardy and E.M. Wright comment as follows: "It is plain, if this be so, that 9 is

n	4	5	6	7	8	9	
$G(n) \leq$	19	41	87	193	425		Hardy & Littlewood (1928)
				137	163		Heilbronn (1936)
16*	23	36		53	73		Davenport (1939, 1942)
		22	34	50	68	87	Narasimhamurti (1941)
		21	31	45	62	82	Thanigasalam (1982)
							Vaughan (1986)

* $G(4) = 16$ exactly.

not the number which is really most significant in the problem. The facts, if they be facts, that just 2 numbers require 9 cubes, and just 15 require 8, are, so to say, arithmetical flukes... The most fundamental and most difficult problem is that of deciding, not how many cubes are required for the representation of all numbers, but how many are required for the representation of all large numbers, that is, of all numbers with some finite number of exceptions".

Thus there is a second, deeper quantity, called $G(n)$, defined to be the smallest number of n th powers needed to represent any whole number, with finitely many exceptions. The "facts, if they be facts," imply that $G(3)$ is 7 (or less) rather than 9.

Because no number of the form $8k+7$ can be a sum of three squares, $G(2)$ is the same as $g(2)$, namely 4. But in general $G(n)$ will be smaller than $g(n)$.

Many authors have studied the functions $g(n)$ and $G(n)$, either for special values of n or for general ones. Thus, in 1942 Linnik showed that $G(3) \leq 7$ (so that the "facts, if they be facts" be facts). Between 1920 and 1928 Hardy and Littlewood developed an approach by way of complex function theory, and proved that $G(n)$ is never larger than $(n-2)2^{n-1}+5$. They improved this to yield $G(4) \leq 19$, $G(5) \leq 41$, $G(6) \leq 87$, $G(7) \leq 193$, $G(8) \leq 425$ and so on. By comparison the best results known by 1930 for $g(n)$ were 37, 58, 478, 3,806 and 31,353, respectively.

Vinogradov introduced powerful methods based on estimating trigonometric sums (a standard approach in analytic number theory but not an easy one to carry through), and in 1936 Heilbronn simplified the method to yield

$$G(n) \leq 6n \log n + [4 + 3 \log(3 + 2/n)]n + 3$$

whence $G(7) \leq 137$, $G(8) \leq 163$. A list of various subsequent improvements is presented in the table (for references see the works cited above). Only for $G(4)$ is the current bound known to be best possible.

Vaughan's methods are related to those of Hardy and Littlewood and Davenport, and involve very careful and delicate analytical estimates. They have applications to other questions in number theory, and may also shed light on Waring's problem for tenth and higher powers.

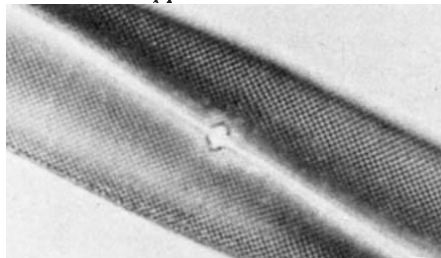
Although number theory may not be

the most directly useful part of mathematics, it exerts an endless fascination because of the gulf between the simplicity of its raw materials and the complexity of its methods. There is some practical payoff too (see, for example, M. Schröder's *Number Theory in Science and Communications* 2nd edn; Springer, New York, 1986). It is remarkable that a problem posed more than two centuries ago, based on simple empirical observations, continues to exercise the ingenuity of number theorists and to generate new mathematics. The Waring experience is far from over. □

Ian Stewart is in the Mathematics Institute, University of Warwick, Coventry CV4 7AL, UK.

Mystery solved

SIR—The miniature crossword that the Churchill Hospital team observed in their chromosome count was probably the remains of a pennate diatom. These organisms produce siliceous frustules which feature the regular pattern seen in their micrograph. The pattern is due to perforations, and if a proportion of them retain air-bubbles during mounting they would convey the crossword appearance, as tiny air-bubbles appear black in mounted



The answer to the crossword question posed in Nature of 25 September?

preparations. The fossilized remains of diatoms — as diatomite — are used as an absorbent (dynamite is composed of nitroglycerine and diatomite, for example). In the laboratory diatomite has been used in filters. None of this should discourage the adventurous wordsmith from writing clues for the puzzle, however, which would be a feat in itself.

BRIAN J. FORD

Mill Park House,
57 Westville Road,
Cardiff CF2 5DF, UK

SIR—Identification of the “mystery object amid the chromosomes” reported by Wolstenholme *et al.*¹ was straightforward. More intriguing was what this object, clearly genetic in nature, could predict about the future of the individual amongst whose chromosomes it was found.

Sadly, the future is not bright. After 47 generations he (or she — only a part of the chromosome spread was visible) turns inescapably into a traffic light. That’s Life².

ANDREW BALL

Biochemistry Department and Biophysics
Laboratory,
University of Wisconsin,
Madison, Wisconsin 53706, USA

1. Wolstenholme, J. *et al.* *Nature* 323, 300 (1986).

2. Berlekamp, E., Conway, J.H. & Guy, R. in *Winning Ways* Vol. 2 (Academic, New York, 1982).

Turing’s fly

SIR—The view that compartmentalization¹ is a crucial event in early *Drosophila* development has frequently been proposed in this journal (see, for example refs 2 and 3). In a recent News and Views article North⁴ reported Meinhardt’s suggestion⁵ that imaginal disks are specified on the embryonic surface at the points of intersection of antero-posterior (A/P)

and dorso-ventral (D/V) compartment boundaries. This suggestion cannot be strictly true as the antennal disk, for example, is specified well before its A/P compartment boundary, which is only established during the second larval instar⁶. No D/V compartment boundary has been demonstrated in the antennal disk and even in the wing⁷ and leg⁷ disks there are no D/V boundaries until late in development.

Whether lineage restrictions analogous to compartment boundaries do exist in the early embryo is a critical question. Clearly, the expression of fate-determining gene products such as those from *Ultra-bithorax* (*Ubx*) and *fushi tarazu* (*ftz*), as revealed by *in situ* hybridization, is restricted to discrete domains. In the case of *Ubx* the domain of expression cannot be compartment-specific because, as predicted by Lewis⁸, it extends across several segments, well into the posterior abdomen⁹. Maximum expression of *Ubx*, however, occurs within a sharply defined region between the middle of the third thoracic and the middle of the first abdominal segments. The boundaries of the domain of maximum expression probably do not represent lineage restrictions because, even by the end of larval development, such A/P restrictions cannot be demonstrated in the larval cuticle¹⁰.

In the case of *ftz*, the domains of expression cannot correspond to a discrete cell lineage because, at the beginning of gastrulation, the *ftz* domains contract from four to three nuclei in width whereas the *ftz*⁺ domains broaden¹¹.

To some extent this argument is semantic: it is clear that there is a set of fate-determining gene products that are expressed within discrete domains. It is these domains of expression, rather than the presence, or absence, of cell-lineage restrictions at their boundaries, which are critical. Where the problem really comes home to roost, however, is with the idea that these domains represent discrete units of pattern formation³. If this were the case, then we have a model that is capable of allocating cells to A or P compartments. However, in both the adult and embryonic ectoderm, the pattern of differentiated cell types indicates that the determination of regional fate is more precise than simply to a half-segmental region. If the fate-determining gene products served only to allocate cells to discrete compartmental or “parasegmental”¹² units, then a separate mechanism would be required to subdivide such units. An alternative possibility is that the discrete domains of these gene products overlap to give each row of cells along the A/P axis of the embryo a unique “positional identity”¹² allowing a much more precise determination of regional fate¹³. There is no requirement for discrete units of pattern formation.

The question posed by Alan Turing, of how the information necessary to describe

an adult organism is generated, is beginning to be answered.

D. GUBB

Department of Genetics,
University of Cambridge,
Cambridge CB2 3EH, UK

1. Garcia-Bellido, A., Ripoll, P. & Morata, G. *Nature* 245, 251-253 (1975).
2. Morata, G. & Lawrence, P.A. *Nature* 255, 614-617 (1975).
3. Martinez-Arias, A. & Lawrence, P.A. *Nature* 313, 639-642 (1985).
4. North, G. *Nature* 322, 404-405 (1986).
5. Meinhardt, H. *Dev. Biol.* 96, 375-385 (1983).
6. Struhl, G. *Nature* 270, 723-725 (1977).
7. Steiner, E. *Wihelm rox Archiv.* 180, 9-30 (1976).
8. Lewis, E.B. *Nature* 276, 565-570 (1978).
9. Akam, M. *EMBO J.* 2, 2075-2084 (1983).
10. Szabad, J., Schubach, T. & Wieschaus, E. *Dev. Biol.* 73, 256-271 (1979).
11. Carroll, S.B. & Scott, M.P. *Cell* 43, 47-57 (1985).
12. Gergen, J.P., Coulter, D. & Wieschaus, E.F. in *Gametogenesis and the Early Embryo* (ed. Gall, J.) (Liss, New York, in the press).
13. Gubb, D. *BioEssays* 2, 27-31 (1985).

Imanishi’s impact in Japan

SIR—Beverly Halstead’s¹ commentary on Kinji Imanishi and anti-darwinian theory has prompted critical replies²⁻⁵ regarding some oddities in his conclusions. Halstead has in fact raised something of a red herring in considering only Imanishi’s anti-darwinism. His impact in Japan does not rest only or mainly on this. The emphasis on discussing Imanishi’s evolutionary theory in the West stems partly from ongoing debates among evolutionists but also from the fact that there is little else written by or about the full extent of Imanishi’s ideas in English. His evolutionary theory was introduced to an English-speaking audience in 1982⁶. The subsequent publication^{7,8} first alerted Halstead to Imanishi’s theory and apparently set the parameters for his investigation.

Halstead states¹ that “... in Japan the average intelligent layman’s understanding of evolution stems in great measure from the writings of Imanishi”. He nowhere indicates that he ever spoke to a sample of such people but given that darwinian theory was introduced to Japan in 1878⁹, that the *Origin* has been translated into Japanese in several editions since 1896^{10,11} and is taught in schools, it is doubtful that they are less cognizant of darwinism than their Western counterparts. In my experience it is Imanishi’s early writings¹² to which the general public refer when asked if they have read him. Those writings contain the kernel of ideas Imanishi later developed into his evolutionary theory, but they are also a combination of a philosophy, personal credo and natural history. His earliest book¹², in particular, evocatively described¹³ as *tamamushi* (referring to an iridescent beetle whose colours change according to the angle from which it is seen), continues to give fresh insights today.

Scientists, too, have derived inspiration from these early books¹⁴. The concept of habitat segregation (*sumiwake*) influenced research by population ecologists

from the early 1950s, especially on dispersal studies in insects^{15,16}. Nor is the idea moribund today^{2,4} and Itô and Iwasa¹⁷ suggest that some valuable criticisms of western theories in ecology may be gleaned from Imanishi.

Imanishi's species-society concept, for which he coined the term *specia*^{7,18} is also important. *Specia* is not the equivalent of a biological species but is a sociological concept that implies a system consisting of member individuals which support it through their interactions^{18,19}. Its influence is most clearly seen in Japanese primatology. Based on the *specia* concept, the study of diachronic and synchronic aspects of social groups and individual life histories of primates, and the focus of social structure and its evolution rather than mechanisms by which individual behaviours are determined²⁰, coupled with Imanishi's advocacy of more analogy with human behaviour^{12,18-20}, led to several differences between Japanese and Western research on primates.

Halstead believes that due to a "virtual absence of debate or intellectual confrontation" in Japan there has been almost no response to Imanishi's evolution theory. In fact, the majority of Japanese scientists simply disagree with his recent popularizations and find these views obscure and untestable. Their silence is hardly surprising. A parallel situation existed in the West in that until the Creationist lobby appeared to be a real threat to freedom of scientific instruction in schools²¹ scientists ignored their anti-evolutionary polemics as being unscientific and not worth commenting on. Instead, Imanishi is widely recognized in Japan for his innovative ideas and pioneering achievements in both biological and social anthropology. He initiated primate behaviour studies in Japan and headed the first African primate expedition. He pioneered research on hunter-gatherer and nomadic pastoralists in East Africa. Many social anthropologists and primatologists have been trained through projects initiated by Imanishi²². Several former students are active in top positions in ethnology, ecology, anthropology and primatology throughout Japan. Some of the unique contributions they have made took root under Imanishi's early influence.

As for Imanishi's evolutionary theory, Halstead's criticism of the culture in which it is embedded is as apparent in his book²³ as in his article¹. His book does not enlarge on a systematic criticism of Imanishi's theory; more simply, it is a diary of his short sojourn in Japan.

PAMELA J. ASQUITH
Department of Anthropology,
University of Alberta,
Edmonton, Alberta T6G 2H4, Canada

2. Sibatani, A. *Nature* **320**, 492 (1986).
3. Sinclair, M. *Nature* **320**, 580 (1986).
4. Millar, C.D., Phillips, N.R. & Lambert, D.M. *Nature* **321**, 475 (1986).
5. Nakahara, H., Sagawa, T. & Fuke, T. *Nature* **321**, 475 (1986).
6. Sibatani, A. *Kinji Imanishi and his anti-selectionism* (Paper presented at the Darwin Conference, London, 5 July, 1982).
7. Sibatani, A. *Rivista di Biologia* **76**, (1), 25-52 (1983).
8. Sibatani, A. *J. Social Biol. Struct.* **6**, 335-343 (1983).
9. Morse, E.S. *Japan Day by Day* (Kobunsha, Tokyo, 1936).
10. Tachibana, S. *Seibutsu-Shigen* (Ichimei Shugenron) (Keizai-zasshi-sha, Tokyo, 1896).
11. Ōsugi, S. *Shu no kigen* (Shinchō-sha, Tokyo, 1914).
12. Imanishi, K. *Seibutsu no Sekai* (Kōbunndo, Tokyo, 1941); *Seibutsu Shakai no Ronri* (Mainichi Shimbunsha, Tokyo, 1949).
13. Takasaka, H. *Seibutsu Kagaku* **36**(2), 94-98 (1984).
14. Suzuki, A. in *Nijusseiki shizen kagaku* (20th century) (ed. Makoto, N.). (Sanseido, Tokyo, 1983). (In Japanese).
15. Solomon, M.E. *A. Rev. Ent.* **2**, 121-142 (1957).
16. Itô, Y. *Kotaijū Seitaigaku no kenkyū* (Research on Population Ecology) **1**, 36-48 (1952) (In Japanese with Eng. summary); *Bull. natl. Inst. agric. Sci. Ser. C*, **11**, 45-130 (1960).
17. Itô, Y. & Iwasa, Y. *Res. on pop. Ecol.* **23**(2), 344-359 (1981).
18. Imanishi, K. *Shizen* **12**(2), 1-9 (1957).
19. Asquith, P. *Proc. tenth Cong. Int. Primatol. Soc.* **3**, (1986); *Shiso* **717**, 36-51 (1984).
20. Itani, J. *Man* **20**(4), 593-611 (1985).
21. Richard, A.F. *Amer. Anthropol.* **83**, 517-533 (1981).
22. Umesao, T. *Senri ethnol. Studies* **3**, iii-ix (1979).
23. Halstead, B. *Kinji Imanishi — The View from the Mountain Top: A Critique of Imanishi's Theory of Evolution*. (Tsukiji Shokan, Tokyo, in the press). (In Japanese).

The northerly extent of Chernobyl contamination

SIR—The radioactivity of snow collected in May 1986 in Svalbard and in Greenland clearly showed contamination of extreme north latitude locations by the Chernobyl accident. Suitably large (several kg) snow samples were subjected to gamma spectrometry, smaller samples were measured by global beta radioactivity (Table 1).

In Svalbard, we observed the first radioactivity as soon as May 4 in a snow precipitation on the Austfonna glacier (79°50' N, 25°30' E); there were further arrivals of radioactive products on this glacier with precipitations of 10, 11, 17, 18

and 25 May. The greatly contaminated deposition of May 10 and 11 may be explained by the high-pressure (1,035 mbar) centred on Novaya Zemlya moving progressively to the East and the low-pressure area spreading from Iceland to the North Pole. On 11 May, a warm front crossed the Austfonna glacier, north east of Svalbard, at 06:00 at a speed of ~ 15 knots, moving first to the north and then towards the north-west, heading for Greenland. It was followed by a cold front going past the Austfonna glacier at ~ 18:00 on 11 May, preceded by strong precipitation.

At Ny-Alesund (78°55' N, 11°56' E) station, the old snow crust sampled on 8 May was not radioactive, but the fresh snow collected on 13 May was contaminated, which shows that the greatest amounts of pollutants have been washed out by the snow precipitation, as has been previously observed for rain¹, and that the dry deposition is at a low level.

The fresh snow samples of May 8 and 12 from 81°36' N and 84°13' N show that the pollution has reached these extreme latitudes and that the calculated concentrations are of the same order as that measured during the high atmospheric nuclear tests in 1963².

The many glaciers present in the Northern Hemisphere constitute an accurate record of isotope release from Chernobyl.

M. POURCHET
J.F. PINGLOT

Laboratoire de Glaciologie et
Géophysique de l'Environnement,
38402 St Martin d'Hères, Cedex, France
J.C. GASCARD
Laboratoire d'Océanographie dynamique
et de climatologie,
Université P. et M. Curie, Paris 6, France

1. Hohenheimer, C. et al. *Nature* **321**, 817 (1986).
2. Holdsworth, G. et al. *Atm. Env.* **18**, 461-466 (1984).

Table 1 Gamma activity, 6-12 May and beta activity on 3 June

		Gamma activity		Beta activity
		Cs ¹³⁴ + Cs ¹³⁷ (Bq kg ⁻¹)	Cs ¹³⁴ /Cs ¹³⁷ ratio (%)	(Bq kg ⁻¹)
Greenland				
Station Nord 81°36' N, 16°40' W	8 May snowfall (13-19 h GMT)	0.38	54	0.38
	12 May snow surface (11 May snow fall)	0.42	47	
Sea Ice Station 84°13' N, 17°50' E	12 May snow surface (11 May snow fall)			1.33
Svalbard				
Austfonna	4 May snow surface (old deposition)	0.02		0.15
Austfonna	4 May snowfall	0.15		0.46
	10-11 May snow fall			8.84
	17-18 May snow fall			1.38
	25 May snow surface (after blowing snow)	0.41	60	1.08
Ny-Alesund	8 May snow surface (old deposition)	0.00		0.05
	13 May snow surface (after snow fall)	0.14		0.15

Data obtained with assistance from ARCTEMIZ, "Femmes pour un pôle", Laboratoire de Glaciologie, M. Riviere (French Embassy) M. Richez and Tromsø Meteorological Station.

The relaxed murderers

Benno Müller-Hill

The Nazi Doctors: Medical Killing and the Psychology of Genocide.

By Robert Jay Lifton. *Basic Books*: 1986. Pp.543. \$19.95.

THE organizers of the mass murder of Jews in Auschwitz and elsewhere were medical doctors. What went on in the heads of these doctors then, and how do they view their actions now? How does a healer turn into a killer?

Robert Jay Lifton provides us with a huge mass of material pertaining to these questions. He studied not only the documents reporting the killings but also spoke to many Nazi doctors who were, to differing degrees, involved in it. To get the view of those who were victims, he also interviewed doctors who had come as prisoners to Auschwitz, including one who had collaborated outright with the murders. To interview the prisoner doctors was painful enough, but to talk with the murderers and their helpers was clearly much worse. Lifton discusses this situation with candour in the Introduction. He used an interpreter in these conversations, which increased the distance but nevertheless allowed the German doctors to speak German and thus express themselves more clearly and fully. He promised anonymity to those who wanted it. Thus some of the prisoner doctors and all of the Nazi doctors remain anonymous.

Lifton describes first the various extermination programmes and then proceeds with the interviews. He concentrates on the "Euthanasia" programme, which was intended to eradicate all mental illness by extermination of the patients. He then moves on to Auschwitz. To understand Auschwitz implies that one wants to understand "a world which is not this world", as one prisoner doctor put it. The coldest, most frightening account and analysis is given by one of the former Nazi doctors. Lifton calls him Ernst B. For anybody who has read about Auschwitz he is easily recognized as Münch, because he is identified as the only camp doctor who refused to select anybody for the gas chamber. Münch describes the conditions at Auschwitz in the following way. There was too little food for the prisoners. The average prisoner died after three months from starvation. The masses of newly arriving Jews were thus directly selected for killing. The doctors' talk centred around technical problems, the most urgent of which was not the killing but the disposal of the corpses. Ethics was a word which did not exist.

Lifton discusses in detail the minds of

Münch, of Mengele and of Wirts, the doctor chiefly responsible for the extermination programme. He then proceeds to examine traits or mental processes common to all of the Nazi doctors, and names two: "doubling" and "numbing". These imply that the typical Nazi doctor appeared as a normal, warm-hearted person with his family, children and friends, but that he had numbed all feelings towards Jews in the camp and had a valueless, scientific and technological view of the realities of Auschwitz. Clearly, many of the Nazi doctors found stability and even consolation in scientific research. Lifton writes here at length about Mengele as the prototype of the "scientifically-relaxed" murderer — "outside of Auschwitz he might have become just a slightly sadistic university professor", is the opinion of one of the prisoner doctors. Also typical is the case of a member of a killing centre of the "Euthanasia" programme, who found consolation in science after a "leading neuropathologist" visited him in his killing centre and explained to him what a wonderful wealth of material was passing through his hands. So he began preparing the interesting brains for the "leading neuropathologist" and even went as a guest research fellow to his Berlin Institute. Again, one of the two men involved in this episode can be recognized quite easily: the eminent research worker from Berlin was the late Professor Hallervorden, former Director of the Max-Planck-Institute for Brain Research.

From passages such as these one gets

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

the disturbing impression that for the pure scientific mind no values exist: the true scientist can be happy even when coming close to mass murder. We hear, for example, about a conversation Lifton had with a psychiatrist who had treated members of the *Einsatzgruppen*. The killers had had nervous breakdowns, anxiety and feelings of guilt for their manual mass-killing of Jewish men, women and children alike. The German psychiatrist recalls how he succeeded in a relatively short time in ridding his patients of their symptoms, particularly the anxiety feelings, and thus was able to send them back to the Soviet Union to kill more Jews. "Did you have anxiety dreams treating these men?" Lifton asked the psychiatrist. "No, I never killed anybody" was the answer.

In the last chapter, Lifton tries to outline the psychology of genocide, that is, the cultural conditions which may lead a people to become organized to commit mass murder. As much as I admire his analysis, I have to say that this is the least successful part of the book. Lifton has hardly read his original sources, almost always quoting them second-hand be they Luther, Goethe, Nietzsche or Max Weber: "Faust quoted as in Pinson" is unfortunately a typical reference in the Notes.

There are just two points where I have disagreements of substance. First, the Jews were not the only "race" on which genocide was planned; the Gypsies were persecuted just as severely. Secondly, I do not share Lifton's opinion that the treatment of "schizoid inmates" with electric shock in Auschwitz was "proper research". Other readers may have other criticisms because a large and detailed work such as this is bound to contain occasional mistakes, misprints and misjudgements. But they cannot detract from the fact that Lifton's book is a monumental achievement, which will for a long time be considered to be the definitive account of the psychology of the murderous Nazi doctors. It should be read carefully and fully — and it makes no easy reading — by any medical doctor or scientist. Above all, the question Lifton asks at the end should be asked and answered by all of us: "can we interrupt the process [of doubling leading to genocide] first by naming it?" □

Benno Müller-Hill is a Professor in the Department of Genetics, University of Cologne, D-5000 Cologne 41, FRG, and author of Deadly Science: The Elimination of Jews, Gypsies and the Mentally Ill, 1933–1945, published in German in 1984 by Rowohlt Taschenbuch Verlag.

• *The Policies of Genocide: Jews and Soviet Prisoners of War in Nazi Germany*, edited by Gerhard Hirschfeld and recently published by Allen & Unwin, is a collection of essays by five German historians on aspects of the systematic killing programme operated by the Third Reich. Price is hbk £18, \$22.95; pbk £5.96, \$9.95.

Blooming thistle

J.V. Golinski

A Hotbed of Genius: The Scottish Enlightenment, 1730–90. Edited by David Daiches, Peter Jones and Jean Jones. *Edinburgh University Press: 1986. Pp.160. Hbk £22.50; pbk £10.95.*

HISTORIANS have devoted much attention to the Scottish Enlightenment. During that intense burst of cultural activity, concentrated in the middle few decades of the eighteenth century, Scottish intellectuals made decisive contributions to fields as diverse as history and sociology, moral philosophy and aesthetics, psychology and epistemology, medicine and chemistry, agriculture and the fine arts. To explain such an intellectual supernova poses a considerable challenge. Was it all a reaction to shrugging off the straitjacket of Calvinist religiosity, or a consequence of the rapid urban and economic development of Scottish society, or a kind of displacement activity for an aristocracy deprived of its governing role after the 1707 Union with England? All these reasons have been suggested, and most probably all of them, and more, must be taken into account for a full explanation.

A Hotbed of Genius does not aim to arbitrate between the accounts of professional historians, but to offer non-specialists an opportunity to savour the climate of the times. To this end, it is extensively and tastefully illustrated, and informally written. David Daiches sets the scene in the first chapter, with a readable review of the Scottish Enlightenment as a whole. The strong interests of the Scottish thinkers in the historical development of their society and the cultural identity of their nation come through clearly; and attention is also devoted to the urban architecture in which they took such pride, and the clubs and societies of which they were so fond. Somewhat more speculatively, Archie Turnbull rounds off the volume with a discussion of possible Scottish influences on the founders of the United States. In between are four studies of leading individual figures of the movement: the philosopher David Hume, the political economist Adam Smith, the chemist Joseph Black and the geologist James Hutton.

Hume and Smith are the two towering geniuses of the Scottish Enlightenment, whose influence is still felt today, although their current reputations do little justice to the range of their concerns. Both men were philosophers in the best sense of the word, concerned with all the moral problems of human behaviour in a changing society, and committed to making pure thought directly relevant to practical action. To perceive the breadth of their

interests is to have one's admiration for them increased, and to begin to see them as men of their time. The articles by Peter Jones and D.D. Raphael have the virtue of helping us to do this.

It is pleasant to see the scientists Black and Hutton getting equivalent recognition; but the essays on them, by R.G.W. Anderson and Jean Jones, while giving well-informed surveys of their careers, do not so convincingly demonstrate their rootedness in their enlightened milieu. The connections which the men of the time saw between Enlightenment culture and the pursuit of chemistry, medicine, agriculture and geology are not explored



From *A Hotbed of Genius*

Enlightened chemist — Joseph Black, the discoverer of carbon dioxide and the principle of latent heat.

here, despite the fact that some recent scholarship has been devoted to illuminating them.

While specialists might carp at details, the general reader is likely to be delighted by this book, not least by its numerous illustrations, many of them in colour. The work of the portraitists Allan Ramsay and Henry Raeburn is rightly exploited to the full by the editors; and the rather weird caricatures of John Kay are also a considerable asset. Architectural and scenic drawings help us to envisage the surroundings in which these men lived and worked. In general, the large investment of the editors in visual material makes for rewarding browsing.

But there is sustenance here for readers as well as browsers. While its illustrations are likely first to draw the eye, the book's text should engage the reader and satisfy much of his curiosity. A coffee-table item perhaps, but with enough intellectual caffeine to stimulate even those to whom the Scottish Enlightenment is familiar ground. □

J.V. Golinski is a Research Fellow at Churchill College, Cambridge CB3 0DS, UK.

Put in the picture

D. R. Higgs

Hemoglobin: Molecular, Genetic and Clinical Aspects. By H. Franklin Bunn and Bernard G. Forget. *Saunders: 1986. Pp. 690. \$85, £98.*

CAPTAIN Marryat may have been a little hasty when he wrote "There is no getting blood out of a turnip". Haemoglobin and related haem proteins have now been identified in a wide variety of animals and more recently, to our surprise, in the leguminous nodules of lupins and in the lowly bacterium *Vitreoscilla*.

Since the relationship between haemoglobin and oxygen transport was established at the turn of the century, this widely distributed molecule has become one of the most intensively studied proteins and with it have been established many of the important landmarks in molecular biology. The abundance of haemoglobin and its messenger RNA in a highly enriched state within red blood cells made it a readily accessible model for studying protein structure, function and synthesis at a time when few other proteins could be examined in this way. Similarly, complementary DNAs and genomic DNA segments encoding haemoglobin genes were amongst the first eukaryotic genes to be characterized. As luck would have it for the molecular biologist, there are also a very large number of naturally occurring mutants of the globin genes which have greatly enhanced our understanding of the haemoglobin molecule and its expression. Unfortunately, these mutants give rise to genetic disorders such as sickle cell disease and thalassaemia that are major causes of morbidity and mortality in many areas of the world.

The enormous literature on haemoglobin is reflected in the fact that *Index Medicus* contains 2,000–3,000 new references on the subject every year. In addition to the traditional categories, studies on haemoglobin genes are playing an increasingly important role in subdisciplines such as molecular evolution, population genetics and anthropology. To marshal this amount of data into good order within a single book is a herculean task, requiring both breadth and depth of knowledge. Franklin Bunn and Bernard Forget make a formidable team to perform this task. Both are medical graduates whose initial interest was kindled by the clinical problems presented by the mutants of haemoglobin. Bunn's special interests tend towards understanding the structure–function relationships of haemoglobin and are complemented by Forget's expert knowledge of the molecular genetics of the globin genes.

The book under review is a follow-up

to a similar, successful edition entitled *Human Hemoglobins*, published nine years ago. The new volume contains all the familiar chapters on haemoglobin structure, function and synthesis, updated as appropriate, but also new chapters featuring animal haemoglobins and their molecular evolution, minor haemoglobin components, methaemoglobin and carboxyhaemoglobin. The sections on the clinical aspects of thalassaemia and sickle cell disease have also been expanded. In addition, "in an attempt to humanize" the material, Bunn and Forget have included a brief history of the major discoveries in the field, and biographical sketches and photographic portraits of a few chosen contemporary investigators. George Orwell observed that "At 50 everyone has the face he deserves" — given the undoubted contribution of these quinquagenarians to the field, I think that they deserved more flattering photographs!

The book is well written and meticulously referenced. Each subject is taken

from simple first principles through to the most complex aspects of our current knowledge, making it of value to novices and *cognoscenti* alike. The illustrations are very clear and the stereodiagrams are outstanding (*sic*). The stereo specs that came with my review copy were essential for my enjoyment of the book, and I feel that with such a relatively expensive publication they should not come as an optional extra.

If I had to pick on anything with which to disagree, it would be in the preface where Bunn and Forget state that our "understanding of hemoglobin is approaching an apogee". On the contrary, I predict that they will be rewriting this book again in ten years time, and including chapters on developmental regulation, tissue-specific expression and maybe some advances towards gene therapy. □

D. R. Higgs is Clinical Research Fellow in the Medical Research Council Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK.

Conscious efforts

D. M. MacKay

The Evolution of the Soul. By Richard Swinburne. *Clarendon: 1986. Pp. 323. £25.*

RICHARD Swinburne is a "dualist" philosopher who believes that we can only make sense of the interdependence of mental states and brain states by supposing that "mental states are states of a soul, a mental substance in interaction with the body". Aware that this is an unfashionable view, he sometimes shows more of the aggressive assurance of a debater than the impartiality of a judge; but it is a pleasure to watch a methodical craftsman's mind at work, and I do not know a better or more persuasive statement of the case for dualist interactionism in the sense he defines. The great merit of the book is that it "takes seriously the fact of human conscious experience, its continuity and its causal efficacy" (p. 3).

Neglect of the primacy of conscious experience has indeed blighted most attempts to see man in purely material terms (the materialist position is remotely plausible only when the materialist is talking about people other than himself). But is dualist interactionism the only reasonable alternative? Among Swinburne's main reasons for thinking so are (a) that sensations are distinct from, not identical with, brain events (Chapter 3); (b) that thoughts are "causally efficacious" (p. 82); (c) that planning with purposes in mind is active and efficacious, conferring evolutionary advantages (Chapter 5); (d) that various imaginary experiments in

which brain hemispheres are transplanted create dilemmas showing that "talk about persons is not analysable as talk about bodies and their parts" (pp. 151, 160); (e) that there is no logical impossibility that a person (suitably defined) could continue to exist without his body; (f) conversely, that knowledge of brains and their states, and knowledge of which mental events were occurring, would not tell you that mental events were states of the same subject (p. 158). Swinburne would "explain" the problem generated by the behaviour of "split-brain" patients as "the problem of discovering the number of souls connected to a given brain" (p. 160).

It has to be said that the notion espoused by Swinburne, that transplantation of an isolated cerebral hemisphere could be sufficient to bring a new person into being, is technically unwarranted. There are structures essential to personality and consciousness in indivisible sub-cerebral brain tissue. But there are many neuroscientists today who would accept and indeed insist upon points such as (a) to (f) as vital and neglected indicators of the complexities of what it means to be human, and as showing the intellectual bankruptcy of traditional materialism.

No doubt, at the cost of generating a host of further problems, to speak of man as made up of two "parts", a body and a soul (Chapter 8), is a possible way of accommodating such facts. It seems a pity, however, that before advancing this as his only possible conclusion Swinburne did not make a more serious effort to reckon with the expansion of our concepts of causality that has come with developments in the theory of information and control. In an information system, we can recognize

"informational" causality as something qualitatively distinct from physical causality, coexisting with the latter and just as efficacious. Roughly speaking, whereas in classical physics the determination of force by force requires a flow of energy, from the standpoint of information theory the determination of form by form requires a flow of information. The two are so different that a flow of information from A to B may require flow of energy from B to A; yet they are totally interdependent and complementary, the one process being embodied in the other. Here we have an irreducible duality, yet no need for dualist interactionism.

If the facts of our conscious experience are related to the physical events in our brains as control by information-flow is related to the physical events that embody it, this would allow (*inter alia*) that sensations can be categorically distinct from and not identical with brain events, that mental activity can efficaciously determine the form of bodily behaviour, that talk about persons is not analysable as (solely) talk about bodies and their parts, and that it is logically possible for a person to exist apart from his present embodiment. It would safeguard the substantial (albeit multidimensional) unity of the human being as we know it, without incurring the scientific and philosophical costs of dualist interactionism. It would allow recognition in human (and animal) nature of a "duality" that has no need of "dualism" in the sense of a "two-part" thought model. (Is the evaluator a "part" separate from the physical works of a thermostat? Is a flame a "part" separate from the gas/air mixture in a burner?)

To call X "part" of Y implies a relation such that a boundary of separation can be drawn between them. Nothing in Swinburne's argument (though we have here only inadequately sampled it) requires or justifies such a step. Minds are not mere computer programs; but the distinction between the equation being solved by a computer (which determines the form of its behaviour), and the physical hardware, points to a kind of conceptual separation quite as radical as that between two "substances", yet one which would not justify any talk of the equation as a separate "part" except in a Pickwickian sense. In this direction, perhaps, lies a more promising way forward? □

D. M. MacKay is Emeritus Professor in the Department of Communication and Neuroscience, University of Keele, Staffordshire ST5 5BG, UK.

• *Nature's* Autumn Books Supplement appears in the issue of 13 November, with contributions from (among others) Philip Kitcher, Lewis Thomas, D.J. Weatherall, Martin J. Klein, Joseph H. Taylor and John C. Marshall. The books reviewed include *The Blind Watchmaker*, *Was Einstein Right?*, *The Harvard Guide to Influential Books and Islands*.

Singular advantage

Clive Ellory

Transport and Diffusion Across Cell Membranes. By Wilfred D. Stein. *Academic:1986. Pp.685. Hbk \$79.50, £67; pbk \$49.95, £42.50.*

TWENTY years on from the first edition, the appearance of a new version of Wilfred Stein's book counts as quite an event. Single-author scientific works are losing out to multi-author volumes, and are now an endangered species, but the present book has all the traditional hallmarks of true authorship — enthusiasm, lucidity and a certain bias or even eccentricity in the selection and treatment of some topics. Stein has enlisted the aid of Bill Lieb for the excellent chapter on simple diffusion across membranes, but otherwise he has taken on the mammoth task of treating membrane transport single-handedly and singlemindedly.

Though there is nothing new in it, the usual introduction to membrane structure and function is sound, while the methodology section is brief but informative. The kinetic approach to behaviour of channels is original and leads directly into the comparison with carriers and what Stein still calls "facilitated diffusion". Here the success of the kinetic analytical approach to a number of transport systems is demonstrated in detail, with examples from glucose, nucleoside, choline and anion transport.

The complexity of the material increases in the fifth chapter, where co-transport systems are considered. As the author admits, it is the topic he found hardest to summarize and it is inevitable that individual readers will have their own list of omissions. Personally I found the lack of any treatment of the Na^+/H^+ exchanger or Na-K-Cl co-transport system (apart from some eccentric references in the table of data) disappointing in view of

their increasing importance in cell physiology. Even the title of this chapter, which refers to the co-transport of two substrates, is misleading because examples of systems carrying three substrates are analysed. The ultimate level of active transport is reached in the final chapter, where nearly 100 pages are devoted to the sodium pump. The enormous literature and byzantine complexities of this system are skilfully summarized and annotated; this part of the book contrasts markedly to recent reviews on the subject, which tend towards the catalogue approach, and it should provide a good source for those seeking explanation and elucidation of mechanisms. Other ATPases are given a less comprehensive treatment. F_0F_1 -ATPases are dealt with sketchily in 14 pages, principally to emphasize the differences with the phosphoenzyme ATPases.

Excellent features of the book are the comprehensive tables and the summaries, and the use of data and figures from original papers. One minor irritation is the poor quality of symbols in the reaction schemes throughout Chapter 6. Another, more serious one, is the index, which is not thorough enough. Inevitably, there are some errors; particularly surprising is the consistent definition of Tl as tellurium! As a member of one of the smallest and most obscure groups working in membrane transport, and specifically identified in the preface as a potentially aggrieved reader, I did not feel my subject (amino acid transport in red cells) had been mistreated. Those in other areas, however, might have a stronger case.

For all the provisos, this book will be essential reading for a large number of scientists. It deserves to be an enormous success. My final impression, having grown up scientifically with its predecessor, was one of slight anticlimax, but Stein's achievement in producing another classic is undeniable. □

Clive Ellory is Reader in the University Laboratory of Physiology, Parks Road, Oxford OX1 3PT, UK.

IMAGE
UNAVAILABLE
FOR
COPYRIGHT
REASONS

Before the telescope — the old observatory of Beijing, equipped with instruments brought to China by the Jesuits in the late 1600s. The picture is taken from Of Stars and Men, the autobiography of Zdeněk Kopal, published by Adam Hilger. Price of the book is £29.50, \$59.

National performance in basic research

D.C. Smith, P.M.D. Collins, D.M. Hicks & S. Wyatt

A newly published report reviews the techniques available for evaluating national performance in basic research and applies these techniques to solid state physics and genetics.

A WEALTH of quantitative data relating to national performance in basic research is contained in two reports just published by the Advisory Board for the Research Councils (ABRC). One of these reports, on the financial inputs to academic research, was discussed in *Nature* last week¹. The second, prepared by the Royal Society's Policy Studies Unit (now the Science and Engineering Policy Studies Unit, sponsored jointly by the Royal Society and the Fellowship of Engineering), reviews a range of the quantitative and semi-quantitative techniques available for evaluating national performance in basic research, and applies them to science as a whole and to case studies in solid-state physics and genetics². It provides for the first time a detailed and objective assessment of the relative performances of a number of countries in these two important areas of research since the early 1970s. This second report is discussed here.

Methodology

The prospect of complementing the more traditional decision-making processes with rigorous quantitative analysis is generating much interest. The Royal Society report shows that, within certain limits, quantitative analysis can indeed produce useful information.

There are two methodologically important parameters to the Royal Society report. First, it deals only with basic research. The output of basic research is typically the published paper, and analysing published papers is a reasonable way of assessing performance. Strategic and applied research are more problematic, because their output has to do with tangible products as well as with papers and open publication may be impeded by commercial or military confidentiality. Second, the report is concerned with national performance rather than, say, departmental or individual performance. There are notable problems in using bibliometric indicators to assess the performance of small groups, because of the nature of the indicators but mainly due to the difficulty in interpreting performances in purely quantitative terms. These problems are alleviated if whole nations are the unit of analysis and if one pays as much attention to shares of world output and to trends over time as to absolute values.

To provide a better test of the methodology, we made case studies in two large

subfields — solid-state physics and genetics — in which UK performance was thought to have been notably dissimilar. We put these in context by studies of the related major fields — physics and biomedical research — and of science as a whole.

We used two types of indicator to assess national performance: indicators of output and of esteem. Output indicators are bibliometric: counts of papers and analyses of citations. Esteem indicators, necessarily less precise, deal with such things as the numbers of scientists deciding to work in particular countries, shares of invited papers at major international conferences and collective opinion about the status of different countries.

It is important to understand the time frames in which the various indicators operate. A certain amount of time must elapse before a change in external factors (for example a change in funding or the opening of some major experimental facility) is reflected in output of papers, and further time elapses before citation scores are affected. The cuts in university funding initiated by the British government in 1981, for example, are unlikely to have had time to influence our bibliometric measures to any great extent. Esteem indicators change at various speeds and may reflect, for example, opinions formed many years earlier or morale in the wake of a recent announcement of cuts in funding. These time lags must be borne in mind when interpreting results from different

indicators.

In the following analysis we shall concentrate mainly on five countries: France, FRG, Japan, the UK and the USA. These are not always the five most active countries in each field, so statements like "the UK ranks second" refer only to these five. In view of the origins of the study, our analysis emphasizes chiefly the UK performance; but our data could equally be used to focus on the performance of the other countries.

Science as a whole

We used the CHI-NSF database³ to count papers by each of the countries from 1973 to 1982 (Fig. 1a). The main features are a slow decline in UK output and a marked increase in Japanese output. In terms of percentage share of world output, the UK moved from 9.2% to 8.3% and Japan from 5.3% to 7.3%. FRG averaged 6.3%, France 5.4% and the USA 37.2%.

Some of these figures have been published before³, but our citation data (Fig 1b), are new. Large-scale citation data usually refer to the number of citations received in a given year by all published papers of a particular country. This indicator has considerable inertia. A more sensitive measure is to count the citations received in a given year by papers published only in that year and the preceding three years, within which time most papers have reached their peak rate of being cited. This procedure identifies about one quar-

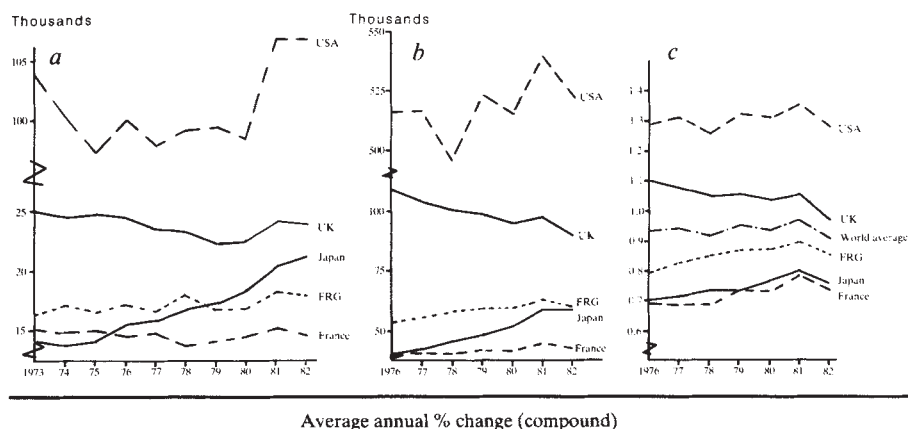


Fig 1 All science (CHI-NSF). *a*, Total annual papers; *b*, citations received each year to papers published that year and the preceding three years; *c*, citations per paper.

ter of all citations, but identifies citation trends more rapidly; it is the normal procedure in citation studies of particular subfields⁴. The UK share of world citations to recent papers fell from 10.9% in 1976 to 8.9% in 1982, in contrast to a steady Japanese rise from 4.1% to 5.8%; FRG averaged 5.8% during this period, France 4.2% and the USA 51.6%.

A further advantage of counting citations to papers published only in a given year and the preceding three years is that, because the number of papers published is known, one can calculate the average number of citations per paper. (For technical reasons explained in the full report, our data on citations per paper are slight overestimates, but the trends in citations per paper for the different countries are more accurate.) Figure 1c shows that individual UK papers in science as a whole are receiving a declining number of citations, whereas German, Japanese and French papers are receiving an increasing number of citations. The decline in total citations received by recent UK papers (Fig. 1b) is thus a result both of the decreasing output and of a decrease in the extent to which each UK paper is cited. Thus the bibliometric indicators from the CHI-NSF database all point to a declining UK performance in science as a whole and to improving performances by FRG and Japan. But by 1982 the UK was still ahead of all countries other than the USA.

One of the indicators of esteem that we used relates to science as a whole: trends in the geographical distribution of Fellows of the Royal Society. Fellows have to be Commonwealth citizens (or 'ordinarily resident' in a Commonwealth country); most are UK citizens. In 1960 13.4% of Fellows were living overseas, 2.8% of them in the USA (and 9.5% in Commonwealth countries other than the UK); by 1986 21.3% were living overseas, 8.1% of them in the USA. At the time of their election, 15.8% of Fellows were living overseas in the three years 1960–62, 3.9% of them in the USA; by 1984–86, those figures had risen to 24.0% and 13.2%. The total numbers involved are fairly small (the Fellowship numbered 603 on 31 December 1960 and 1,022 on 31 July 1986), but they do suggest that a gradually increasing number of the best UK scientists are choosing to work outside the UK.

Physics

Figure 2 presents data for physics from the CHI-NSF database analogous to those given in Fig. 1 for science as a whole. On all three indicators — output of papers, citations to recent papers and citations per paper — a decline in UK performance is evident. The country rankings, however, are different. The UK share of world output of physics papers, as measured by the CHI-NSF database, declined from 7.7% to 6.1% during 1973–82: a faster rate of

decline than for science as a whole and a share of world output that is on average some 25% below the UK share of all science. The UK share of citations to recent physics papers fell from 8.1% in 1976 to 6.3% in 1982, similar to the rate of decline of its share of all science and again an average of some 25% below it. On both indicators Japan and FRG overtook the UK. On citations per paper the UK declined more slowly, but here too there were marked improvements in the German and Japanese performances. On all three indicators France closed the gap on the UK without, by 1982, overtaking it.

For dealing with particular fields of science, numerous other databases can be used. These tend to be constructed for information retrieval rather than policy analysis (as was the *Science Citation Index* itself), but many can be turned to bibliometric ends. For evaluating performance in physics generally and solid-state physics in particular, we made extensive use of *Physics Abstracts* (PA). One notable feature of PA is that it allows one to separate basic research papers from other types of publication and to distinguish experimental from theoretical papers.

Figure 3 shows output of physics papers according to PA, by country and year. This confirms the picture from the CHI-NSF database, with Germany and Japan moving ahead of the UK. The UK averaged 6.3% of world output of both experimental and theoretical papers during 1977–84; Japan, interestingly, had a much higher share of experimental (average 9.6%) than theoretical (average 6.2%) papers and expanded more rapidly in the former than the latter. The average German share of experimental papers (6.7%) was a little better than its share of theoretical papers (6.1%); France too, did better in experimental (5.4% of world output)

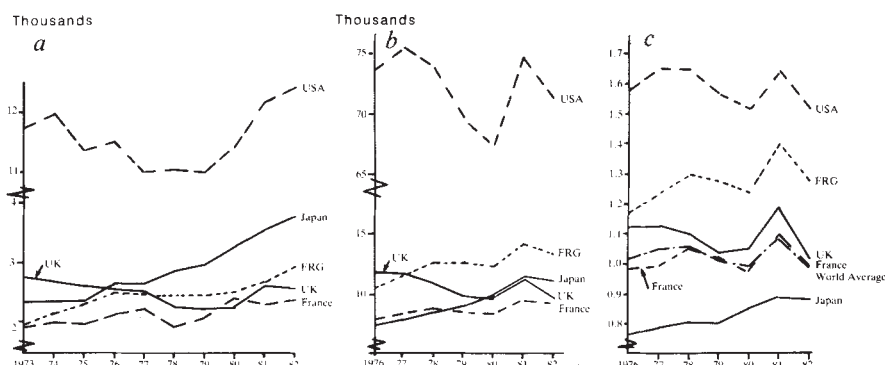
than theoretical (4.1%) papers. The US share of papers in the physics field of the CHI-NSF database averaged 31.1% over 1973–82, as opposed to 37.2% for science as a whole; this relatively poorer performance is evident also in PA, where the United States had only 26.1% of experimental and 26.5% of theoretical papers over 1977–84.

Both the CHI-NSF and the PA databases show that UK performance in physics, by remaining roughly constant in volume terms, has failed to match the growth shown by other countries, is falling behind not only the USA but also Japan and FRG and is being caught by France. Moreover, the UK world share of papers in physics is both well below its world share in all science and declining at a faster rate.

Solid-state physics

Output of papers in the PA category of 'condensed matter physics' is shown in Fig. 4. This is similar to Fig. 3 for all physics, except that France here is ahead of the UK on the experimental side. The UK share declined from 6.4% in 1977 to 5.5% in 1984 (experimental) and from 7.5% to 5.9% (theoretical). The Japanese share of experimental papers in condensed matter physics, averaging 11.8%, was notably higher than its share of experimental papers in all physics combined; its share of theoretical papers was closer to its all physics value. The USA, on the other hand, had relatively lower shares in condensed matter physics than in all physics, at 19.7% (experimental) and 21.8% (theoretical).

We carried out detailed manual citation studies in three areas of solid state physics: spin glass, quantum fluids and EXAFS (extended X-ray absorption fine structure). These were chosen to represent, respectively, a new area, a well estab-



Average annual % change (compound)

	France	FRG	Japan	UK	USA	World average
Papers	2.2 ± 0.7	3.2 ± 0.7	5.7 ± 0.6	-1.3 ± 0.7	0.3 ± 0.3	1.6 ± 0.3
Citations	2.3 ± 1.7	3.8 ± 1.4	7.5 ± 1.7	-2.8 ± 1.5	-0.7 ± 0.6	1.0 ± 0.4
Citations per paper	-0.6 ± 0.9	1.6 ± 0.8	1.8 ± 1.0	-0.8 ± 0.8	-0.8 ± 0.7	-0.3 ± 0.9

Fig. 2 All physics (CHI-NSF). a, Total annual papers; b, citations received each year to papers published that year and the preceding three years; c, citations per paper.

lished area and an area requiring access to a major experimental facility. In experimental spin glass research the UK has lost an early lead and in recent years has produced fewer papers than France, FRG or Japan. It similarly fell behind these three countries in its share of total citations, but remained ahead of West Germany and Japan in citations per paper. Overall, the country's total impact on the discipline is now well behind that of France, FRG and Japan (and USA). The UK performance in theoretical spin glass research, by contrast, has been rather stronger, and in terms of overall impact is second only to the USA.

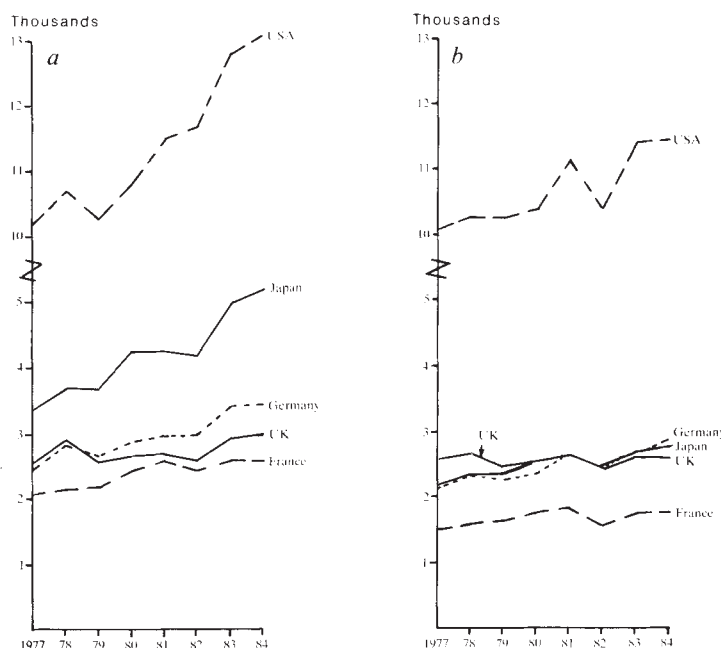
With quantum fluids we found few trends in output of papers during 1971-84. On the experimental side, the UK was second to the USA, with an average of 9.5% of output, and was also second in share of citations. However, in citations per paper the UK came below France and FRG, markedly so in recent years. On the theoretical side the UK was behind both USA and Japan in terms of output and behind also FRG and France in citations.

In EXAFS the UK share of papers during 1974-84, at 7.3%, was below USA, Japan and France and above FRG. The UK share of citations, however, (also 7.3%) was ahead of all countries except the USA, especially in the more recent years. This was also true for citations per paper and numbers of highly cited papers.

These three detailed studies present a complex picture of strengths and weaknesses. The overall impression is that the UK performance in solid state physics, as in physics generally, has been below its performance in science as a whole and is rapidly falling behind that of its major scientific competitors.

For indicators of esteem, we looked at UK participation in the International Conferences on Low Temperature Physics. The number of UK papers at these conferences has remained roughly constant since the mid 1950s, at around 38 per conference. The UK share of papers, however, has dropped from 29% in 1955 and 19% in 1957 to under 5% in 1984. The UK share of specially invited papers has also fallen, though the actual numbers here are too small to permit much analysis. The indicator as a whole corroborates the message of the bibliometric data.

A questionnaire survey of Fellows of the Royal Society with interests in solid-state physics found a rapid advance of FRG and Japan from positions of esteem well behind the UK 25 years ago to virtual parity now. France also appears to be closing the gap, though more slowly. The USA has moved from somewhat ahead of the UK in esteem terms to well ahead. These findings are consistent with our output data described above — if anything, they overestimate the position of the UK relative to FRG and Japan.



	Average annual % change (compound)					
	France	FRG	Japan	UK	USA	World average
Experimental	3.4 ± 1.1	4.3 ± 1.0	5.8 ± 0.8	1.4 ± 1.0	3.7 ± 0.5	3.6 ± 0.4
Theoretical	1.7 ± 1.1	3.5 ± 0.9	2.9 ± 0.9	-0.2 ± 0.9	1.9 ± 0.4	2.0 ± 0.4

Fig. 3 All physics (*Physics Abstracts*). a, Experimental papers; b, theoretical papers.

One other characteristic of UK physics can be discerned from our bibliometric data. We produced, from *PA*, data on shares of world output for physics as a whole and for 19 subfields within physics, giving, with an experimental/theoretical subdivision, 38 sets of data. It is interesting to compare a country's share of world output in each subfield with its share of world output in the whole field and to count how often the former differs from the latter by, say, at least 25%. A score of zero would indicate a fairly even distribution of effort among all branches of physics, a score of 38 a highly selective distribution. The scores we found were: USA, 12; UK, 14; Germany, 15; France, 18; Italy, 21; Netherlands, 22; Japan, 24. What one cannot discern from these data *per se* is whether they are the results of deliberate national policy decisions or simply natural differences between countries with relatively stable or rapidly expanding programmes in physics research.

Genetics

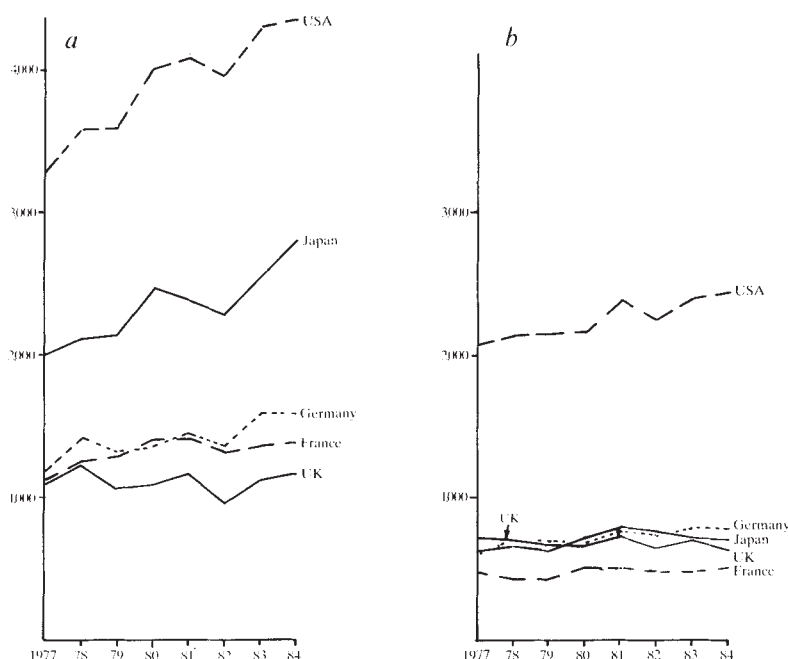
Bibliometric analysis of genetics proved technically more difficult than that of solid-state physics. The CHI-NSF database has two main fields covering the life sciences: biology and biomedical research. Genetics forms part of the latter. The UK output of biomedical papers was fairly constant during 1973-82, and ahead of all countries except the USA; but the UK share of world output of biomedical

papers fell from 9.4% to 8.4% (figures very close to the UK share of output in all science) in the wake of expanding output from the USA and Japan. The UK share of citations to recent biomedical papers was likewise very close to its performance in science as a whole, falling from 10.9% in 1976 to 8.8% in 1982. UK performance in the field of biology, however, was stronger, with an average 10.2% of world output and a share of world citations to recent biology papers that fell from 14.0% in 1976 to 11.7% in 1982.

We used *Biological Abstracts* to measure output in genetics, though for technical reasons only for the period 1980-84. We found no particular time trends. For the period as a whole the UK had 6.9% of world output, behind the USA (38.9%) and Japan (7.0%) but ahead of France (4.6%) and FRG (4.2%).

Against this general background we carried out detailed manual studies of two areas within genetics: the rapidly developing area of genetic instability and transposable genetic elements, and the well established area of *Drosophila* genetics. In the former, the UK had 8.2% of papers during 1970-84, similar to FRG (8.5%) and well ahead of Japan (4.2%) and France (2.8%). The USA had 51.2% of papers. There were no clear time trends. In terms of total citations and citations per paper, the UK was significantly ahead not only of Japan and France but also of FRG.

In *Drosophila* genetics we again found



Average annual % change (compound)

	France	FRG	Japan	UK	USA	World average
Experimental	2.2 ± 1.1	3.5 ± 1.0	4.2 ± 0.8	-0.2 ± 1.1	3.8 ± 0.6	3.1 ± 0.6
Theoretical	2.0 ± 1.1	3.0 ± 0.9	2.4 ± 0.9	-0.9 ± 0.9	2.3 ± 0.5	2.2 ± 0.8

Fig. 4 Condensed matter physics (*Physics Abstracts*). a, Experimental papers; b, theoretical papers

no clear time trends in output of papers during 1974–83. The UK share averaged 6.0% of world output, ahead of France (4.8%), FRG (4.7%) and Japan (4.3%). The USA had 39.9% of world output, though it also had 65.6% of total citations. In total citations and citations per paper, the most striking trend was a marked decline by France; the UK was on average ahead of FRG, France and Japan (in that order).

One of the features of these two studies was that they highlighted the importance to each research area of Switzerland and, to a lesser extent, Belgium and The Netherlands. Switzerland has had a larger impact than either Japan or France in each of these areas.

We performed a rather different sort of citation analysis by counting citations in patents to journal papers to examine the impact of the basic research. We found that a group of 366 genetics patents granted in the USA during 1980–85 contained 3,238 citations to 1,539 different papers in the basic research journals; half the cited papers were less than five years old. Analysis of these citations showed that the UK produced 9.4% of the cited papers, compared with the USA at 63.4%, Japan and FRG at 4.2% each and France at 3.2%. UK papers gained 11.9% of the total citations, compared with the USA at 66.5%, FRG 3.8%, Japan 3.7% and France 2.5%. There were no particular time trends in the papers being cited:

recent UK papers were as likely to be cited as older papers.

Our three detailed citation studies in genetics thus corroborated the CHI-NSF citation data for all biomedical research in showing the UK substantially ahead of FRG, Japan and France. But they did not confirm the impression that the UK's relative position was steadily weakening during the past decade. It would thus appear that genetics is not a party to the declining trends shown by the UK for biomedical research. Indeed, the average UK share of papers in the CHI-NSF subfield 'genetics and heredity' during 1973–82, at 10.2%, is noticeably higher than the corresponding figure of 8.9% for UK papers in the whole field of biomedical research; for all other countries in our set the reverse relation applies. The bibliometric data thus indicate that, in terms of outputs, UK genetics has maintained a position generally ahead of FRG, Japan and France, though behind the USA.

For an indicator of esteem we looked at trends in participation in the Cold Spring Harbor Symposia on Quantitative Biology. For the period 1961–84, we found that 70.0% of invited speakers were from the USA, 7.0% from the UK, 4.8% FRG, 4.1% France and 2.8% Japan. In the more recent years, 1973–84, the shares were 67.9% USA, 7.6% UK, 5.8% FRG, 2.9% France and 2.9% Japan. Apart from a slightly weaker position for Japan, these figures confirm the picture given by the

bibliometric indicators.

A questionnaire survey of senior geneticists in the UK and overseas pointed to a marked decline in the perceived status of UK genetics over the past 25 years. Nevertheless, only the USA was generally thought to be ahead of the UK (by a very long way). The extent to which esteem for German, Japanese and French genetics was thought to have risen relative to UK genetics is not wholly matched by overall trends in our other indicators. It may be that this esteem indicator is influenced by perceptions of the changes in funding and opportunities for scientific research in the United Kingdom that have occurred since the mid 1970s and particularly since 1981.

Conclusion

The Royal Society report contains full details of the various methodologies we used and very extensive sets of tables with bibliometric and other data. The main findings are that in basic research in science as a whole, UK performance has deteriorated since the early 1970s both in absolute terms and relative to other countries, notably FRG and Japan. The UK is, nevertheless, still second only to the USA among the countries we analysed. The same conclusions apply to biomedical research, though in genetics the UK seems to have done somewhat better, maintaining its position more clearly ahead of FRG, Japan and France. But in physics as a whole and in solid-state physics, UK performance has been markedly below its average for all science: since the early 1970s the UK has been overtaken by both FRG and Japan, though it remains ahead of France.

We believe that the quality and usefulness of the data we obtained justify the attempt to undertake quantitative assessments of national performance in basic research. Such assessments constitute a source of hard information for policy-makers, particularly valuable at a time when policy decisions are rapidly becoming more complex and more serious in their long-term consequences.

This study was supported by grants from the Advisory Board for the Research Councils and the Economic and Social Research Council. □

Sir David Smith, secretary of the Royal Society, is chairman of the Steering Group for the Royal Society/Fellowship of Engineering Science and Engineering Policy Studies Unit. Peter Collins is head of the unit. Diana Hicks (now at the Science Policy Research Unit, University of Sussex) and Suzanne Wyatt worked as research assistants on this study.

1. Irvine, J. & Martin, B.R. *Nature* **323**, 591–594 (1986).
2. *Evaluation of National Performance in Basic research. A Review of Techniques for Evaluating National Performance in Basic research, with Case Studies in Genetics and Solid State Physics* (ABRC Science Policy Studies No 1 Advisory Board for the Research Councils, London, 1986).
3. Irvine, J., Martin, B., Peacock, T. & Turner, R. *Nature* **316**, 587–590 (1985).
4. Martin, B.R. & Irvine, J. *Research Policy* **12**, 61–90 (1983).

Stable and metastable metal surfaces in heterogeneous catalysis

M. S. Spencer

ICI PLC, Agricultural Division, Billingham, Cleveland TS23 1LB, UK

Most supported metal catalysts are remarkably consistent in performance, although they are intrinsically unstable. Their success in catalysing 'structure-sensitive' reactions implies that the metal crystallites of catalysts reach relatively stable, quasi-equilibrium states under most reaction conditions. The variability found in some systems is attributable, at least in part, to the presence of metastable catalyst states.

MANY heterogeneous catalysts show a remarkable consistency of performance, despite the instability inherent in an assembly of small particles. In catalysts which consist of microcrystallites of metal on a refractory support, the metal particles are frequently <10 nm in diameter, and are sometimes as small as ~1 nm. Widespread studies^{1,2} with a new silica-supported platinum catalyst, EUROPT-1, have shown excellent consistency. Various reactions of hydrocarbons and carbon monoxide³ have very similar values of turnover frequency (=molecules of product/(number of catalyst sites × time)) on large single crystals and small (~1 nm) supported crystallites. But perhaps the most extensive demonstration of the consistency of heterogeneous catalysts has been their successful industrial use for decades. Some catalytic reactions ('facile' or structure-insensitive^{4,5}) are relatively insensitive to the state of the catalyst surface; in these cases, the independence of catalyst performance on catalyst preparation or activation is not surprising. However, the same phenomenon is seen in many reactions, such as hydrocarbon hydrogenolysis^{4,5} and ammonia synthesis^{6,7}, which are structure-sensitive or 'demanding'. This leads to the conclusion that the metal surfaces promoting the catalysis in many reactions are consistent, for given reaction conditions, both in crystal face and in micro-geometry (steps, defects and so forth). Although it had been thought⁸ that microcrystallites showed different geometries from bulk metals, it now seems probable that bulk structures (such as the face-centred cubic structure) hold down to very small crystallite sizes⁹. In contrast to this pattern, some catalytic systems (considered below) display variable activities, and even different products from the same reagents, when investigated by different workers.

Any working catalyst must be in a (more or less) steady state. For the present purpose we will define two types of steady state in which supported metal catalysts can occur:

(1) Catalysts with equilibrated metal surfaces (EMS). In these catalysts the agglomeration or fusion of the metal crystallites is slow (relative to the timescale of the catalysed reaction), but the surface (that is, the top atomic layer or layers) of each metal crystallite is sufficiently mobile to reach a quasi-equilibrium state. This is the normal condition of most, possibly all, industrial catalysts and of many experimental catalysts. Bond¹⁰ has suggested that the insensitivity to structure of many catalysed reactions could be attributed to surface mobility. As the sintering of metal crystallites in these catalysts is limited by support properties¹¹, surface mobility does not necessarily lead to the loss of active metal surface by agglomeration.

(2) Catalysts with metastable metal surface (MMS). In contrast to EMS, the metal surfaces of these catalysts are not equilibrated because the timescale of mobility of the surface atoms is greater than that of catalyst use. The surface states of the metal crystallites (crystal faces, edges, defects) in these catalysts depend on catalyst history as well as on energetic

considerations. The rates of some catalysed reactions (such as ammonia synthesis on iron⁶ or rhenium¹²), can vary by several orders of magnitude between different crystal faces, so variable catalyst performances would be expected with MMS.

We first examine results from surface physics which show that surface mobility is to be expected under the conditions generally required to get measurable rates of catalysed reactions. Several catalytic systems where MMS may be significant are then reviewed.

Conditions for catalysts with EMS

Surface mobility. Atoms in the surface of a metal move readily under various conditions even when the bulk remains essentially unchanged. Surface atoms have lower coordination numbers than bulk atoms, and thermally activated surface rearrangement is possible. Furthermore, the energetic changes following the bonding of adsorbed species (adsorptive reconstruction) can also cause surface changes. The extent of mobility required to modify the crystal faces of a 5-nm crystallite is very much less than that needed to produce visible changes in a macroscopic crystal. Thermal energy and adsorptive reconstruction are the common causes of surface mobility leading to EMS.

The surfaces of solids become mobile due to thermal energy at temperatures well below the melting point¹³. Facile surface diffusion is seen at much lower temperatures than surface melting¹⁴. Most experimental work on metal surface diffusion has been done at temperatures either close to the melting point or much lower than those of catalytic interest; however, Schrammen and Holz¹⁵ have investigated the surface self-diffusion of nickel atoms on the Ni(100) plane over a wide temperature range. They found that the crystal surface reached a common equilibrated state, dependent on the initial coverage of Ni adatoms, between 500 and 620 K. This corresponds to values of T/T_m (where T_m is the melting temperature) of 0.29 and 0.36 and it agrees well with older work (for example, ref. 16) which indicated the onset of surface mobility at a temperature of $\sim 1/3 T_m$, often called the 'Huttig temperature'. This seems to be generally valid for transition metals. Values for some metals of catalytic interest are listed in Table 1. Any catalyst working above the Huttig temperature of its metal component will probably have EMS, regardless of the reaction catalysed. Catalysts pre-reduced above the Huttig temperature, before use at lower temperatures (for example, nickel/silica catalysts for organic reductions), are also likely to have EMS. Schrammen and Holz¹⁵ have also shown that isolated Ni adatoms, often shown in models of catalyst surfaces, will not be found on Ni(100) above 300 K. The atomic motion in ultra-fine gold particles (containing ~500 atoms), directly observed¹⁷ by electron microscopy, indicates the behaviour to be expected of metal crystallites in catalysts. Thus most, if not all, supported metal catalysts above the Huttig temperature of the metal microcrystallites contain

Table 1 Huttig temperatures (calculated as $1/3 T_m$) for some common catalytic metals

Metal	Huttig temperature (K)
Fe	600
Co	590
Ni	570
Cu	450
Ru	840
Rh	750
Pd	610
Re	1,150
Os	1,090
Ir	890
Pt	680

crystallites which are equilibrated, with respect to both exposed crystal faces and defect concentrations.

Adsorptive reconstruction of crystal faces. The Huttig temperature applies to a solid *in vacuo*. Surface changes can be induced at lower temperature by adsorbed species, including intermediates in catalytic processes. The visible recrystallization of platinum wires catalysing the selective oxidation of ammonia at high temperatures was observed early on¹⁸, and similar surface changes in platinum wires have been found¹⁹ in the catalytic oxidation of CO below 600 K. Surface physics' investigations have shown that adsorbate-induced reconstruction of transition metal surfaces, on a microscopic scale, is a common phenomenon. Examples of catalytic relevance include O atoms on the Cu(110) and (111) faces²⁰⁻²², H atoms on the Ni (100) and (110) faces^{23,24}, S atoms on the Fe(110) face²⁵ and N atoms on the Fe(111) face²⁶, although surface reconstruction does not occur in some similar systems, such as N atoms on the Fe(100) face²⁶. As chemisorption of some intermediates is an essential part of heterogeneous catalysis (as the means of decreasing the activation energy of the reaction), adsorptive reconstruction of the catalyst surface is always possible. The heat evolved in catalysed reactions (even overall endothermic reactions probably have exothermic steps) should also promote restructuring below the Huttig temperature. The etching of platinum foil which occurs in catalysed ethene oxidation has been shown²⁷ to be due to the catalytic reaction itself and not to the sum of the etching effects of the individual reactant and product molecules.

There are two consequences of adsorptive reconstruction for catalysis on supported metals. First, when the catalyst surface reaches a quasi-equilibrium state, this state is dependent on reaction conditions. Second, surface reconstruction provides a mechanism for reaching an EMS below the Huttig temperature. Reconstruction may occur during catalyst activation as well as during reaction: the reduction of an oxide precursor is a common method of activation, so an enhanced surface mobility due to adsorbed oxygen on metal crystallites may be found in the later stages of reduction.

Reactions on catalysts with EMS

We consider ammonia synthesis as a typical example of heterogeneous catalysis. The reaction takes place on iron catalysts in the range 600–800 K, above the Huttig temperature for iron. Moreover, the crystal face probably responsible for catalysis⁷, Fe(111), reconstructs with nitrogen chemisorption. Thus, the active phase of this catalyst probably has EMS. Similar considerations apply for many other catalytic reactions, either to reaction conditions or to catalyst-activation conditions (such as pre-reduction of nickel catalysts for liquid-phase hydrogenations). Some examples are given in Table 2.

The EUROPT-1 catalyst, described above as a consistent catalyst, was pre-reduced at 673 K after preparation. This is

Table 2 Reactions catalysed by metal catalysts (taken from ref. 52) in which the active metal phase probably has equilibrated metal surfaces (EMS)

Reaction	Active metal phase
Catalytic re-forming	Pt
Methane/steam re-forming	Ni
Ammonia synthesis	Fe
Water-gas shift reaction	Cu
Methanol synthesis	Cu
Fischer-Tropsch synthesis	Fe
Liquid-phase organic hydrogenations	Ni
Methanol oxidation to formaldehyde	Ag
Ethylene oxide synthesis	Ag
Ammonia oxidation to nitric oxide	Pt
Methane ammoxidation to hydrogen cyanide	Pt

likely to have been sufficient to generate equilibrated platinum crystallites (see Table 1).

Conditions for catalysts with MMS

There are three requirements for catalysts with MMS, covering both preparation and use. (1) The temperature must be below the Huttig temperature of the metal. This is an approximate constraint, partly because the Huttig temperature is only a broad indication of the onset of surface mobility and partly because of the effects of refractory materials (see (3) below). (2) The catalyst must be free of reconstructing adsorbates. (3) Suitable refractory materials must be present, to prevent the sintering or reconstruction of metal crystallites. This is more important here than for catalysts with EMS because the more energetic crystallites with MMS have a greater driving force for physical change. However, the effects of high-melting-point impurities on surface structure can be complex. Bettler *et al.*²⁸ found that carbon or silicon caused the re-structuring of several transition metal surfaces, but also increased the activation energy for surface self-diffusion.

Catalytic systems with possible MMS

Various catalytic systems fulfil the requirements for MMS, and in some cases variable performance has been attributed to changes in metal crystallite morphology. Some workers have proposed other causes, such as geometric and electronic effects of promoters and supports. Although there is clear evidence that these effects are important, the role of MMS has sometimes been ignored. As MMS have surface atoms in wider ranges of surface coordination than EMS, the consequences of MMS are likely to be more evident in demanding than in facile reactions.

Most nickel catalysts require temperatures above 570 K, the Huttig temperature, either for reduction of the metal precursor (when oxygen-induced reconstruction may also be significant) or for practical rates of reaction. Thus, most nickel catalysts have EMS, but some examples with MMS have been reported. Kuijpers *et al.*²⁹, in a study of carbon deposition from CH₄ and CO on a silica-supported nickel catalyst, found that catalyst activity was increased by carburization-decarburization cycles. They attributed this increase to a very small number of strongly bound carbon atoms (remaining in the surface after decarburization) which inhibited the relaxation of the nickel surface to its original, less active state. Klabunde *et al.*³⁰ have prepared nickel and other transition-metal catalysts from metal atoms in hydrocarbon solvents at low temperatures, that is, in conditions for avoiding EMS. These catalysts, which also contained carbonaceous residues, showed exceptional activity, probably attributable to MMS.

There is now strong evidence³¹ that copper metal is the active phase in the Cu/ZnO catalysts used for methanol synthesis. Under normal reaction conditions (500–580 K, with a partially

oxidized copper surface³¹) the copper crystallites have EMS. Similar catalysts³² prepared by a low-temperature route (Raney copper) show comparable activities and selectivities. In contrast, copper/thoria catalysts made from copper/thorium alloys were found³³ to be both more active and less selective (methane by-product formation) than conventional methanol synthesis catalysts. The method of preparation gives very close interaction between metal and support, probably stabilizing MMS on the copper phase. The use of CO/H₂, without CO₂, as the reactant minimized surface oxidation and reconstruction during methanol synthesis, although water, the co-product of methane formation, allowed some oxidation. Pritchard *et al.*³⁴ found from the infrared spectra of chemisorbed CO that copper supported on magnesia exposed mainly (100) and (111) faces, whereas on glass, silica or alumina substrates the copper surface was composed of high-index faces.

Palladium under similar reaction conditions is below its Huttig temperature. Corma *et al.*^{35,36} have deduced from chemisorption measurements that different preparations of supported palladium catalysts have the crystal faces Pd(100), Pd(110) and Pd(111) in different proportions. The reactions of CO/H₂ mixtures on supported palladium catalysts have shown³⁷ marked variations, both in activity and in selectivity to methanol or methane. While it is clear that basicity of the support and oxidation state of the palladium are relevant factors³⁷, as is Pd crystallite size³⁸; it is also possible that different catalysts contain different MMS of palladium. Hicks and Bell³⁹ found that the Pd morphology rather than Pd dispersion affected the specific activity of Pd/La₂O₃ catalysts for methanol synthesis. Carbon monoxide dissociation on the Fe(110) crystal face is significantly enhanced⁴⁰ by slight sputtering, showing that differences in catalytic performance between EMS and MMS in synthesis gas reactions can be expected.

As discussed above, the conventional iron catalyst used for ammonia synthesis operates with EMS. Under the same conditions ruthenium is below the Huttig temperature, so MMS may be present. Supported ruthenium catalysts are effective for ammonia synthesis, but the activity is strongly dependent on support and promoters^{41,42}. This variability has been attributed^{41,42} to electron transfer and compound formation, although MMS may also contribute. Unlike the Raney copper catalysts in methanol synthesis, Raney ruthenium catalysts have a higher specific activity⁴³ than pure ruthenium, prepared by the reduction of RuCl₂·H₂O at 723 K. These differences in preparative technique would be expected to give different ruthenium crystallites, which would equilibrate only slowly.

Catalytic effects which may be due to the variability of MMS have been observed in other platinum-group metal catalysts.

Burch and Garba⁴⁴ found that the activity of Pt/Al₂O₃ in the hydrogenolysis of pentane depended on platinum surface morphology. In the hydrogenolysis of methylcyclopentane⁴⁵, the specific activity of rhodium catalysts varies with both support and dispersion of rhodium: small rhodium particles were more active than large particles when the support was silica but the reverse was true on alumina. These variations were interpreted as being derived from differences in the surface structure of the small rhodium particles. Variation of surface structure, and consequent variation in reaction rate, were found⁴⁶ in oscillations during the catalytic oxidation of carbon monoxide on Pt(100).

Strong metal-support interaction (SMSI)

Tauster *et al.*⁴⁷ discovered that high-temperature reduction of Pt/TiO₂ and similar catalysts greatly decreased chemisorption and activity in many catalytic reactions—an effect they called strong metal-support interaction (SMSI). It now appears⁴⁸ that the phenomenon is due to the formation of a monolayer (or less) of a reduced titanium oxide, TiO_x, on the surface of the platinum crystallite. However, the rates of a few catalytic reactions, such as the hydrogenolysis of carbon monoxide to methane, are increased by the previous formation of the SMSI state. While this may be due in part to oxidation of the catalyst by the water co-product⁴⁸, a further explanation is required, especially as the effect has now been seen in benzene hydrogenation⁴⁹. This high activity is attributed^{49,50} to some special sites at the interface between support and metal. These can be seen as metastable surface structures stabilized by small amounts of oxide on the metal surface, analogous to the role of traces of carbon in the promoting MMS in nickel catalysts²⁹.

Non-metallic catalysts

The extension of these ideas to non-metallic catalysts is clear in principle, but data on surface mobility and reconstruction are largely lacking. An indication of the phenomena involved was shown by Greger *et al.*⁵¹, who found non-equilibrium phase transitions with hysteresis (CuO ↔ Cu₂O) in copper oxide/alumina catalysts during the oxidation of propene.

Conclusions

The metal crystallites in supported metal catalysts have relatively mobile surfaces under most catalytic reaction conditions. This leads to the attainment of an equilibrated surface, which is the main cause of catalyst reproducibility. When reaction conditions do not permit the formation of equilibrated surfaces, various metastable states of variable catalytic activity can be formed, depending on catalyst history.

- Coenen, J. W. E. & Wells, P. B. in *Preparation of Catalysts III* (eds Poncelet, G., Grange, P. & Jacobs, P. A.) 801–814 (Elsevier, Amsterdam, 1983).
- Bond, G. C. & Yide, X. *JCS Faraday Trans. 1* **80**, 969–980 (1984).
- Boudart, M. & Djega-Mariadasson, G. *Kinetics of Heterogeneous Catalytic Reactions*, 157 (Princeton University Press, 1984).
- Boudart, M. in *Proc. 6th int. Congr. Catal. Vol. 1* (eds Bond, G. C., Wells, P. B. & Tompkins, F. C.) 1–9 (Chemical Society, London, 1977).
- Boudart, M. & McDonald, M. A. *J. phys. Chem.* **88**, 2185–2195 (1984).
- Spencer, N. D., Schoonmaker, R. C. & Somorjai, G. A. *J. Catal.* **74**, 129–135 (1982).
- Boudart, M. & Löffler, D. G. *J. phys. Chem.* **88**, 5763 (1984).
- Burton, J. J. *Cat. Rev. Sci. Engng* **9**, 209–222 (1974).
- Joyner, R. W. & Meehan, P. *Vacuum* **33**, 691–696 (1983).
- Bond, G. C. in *The Physical Basis for Heterogeneous Catalysis* (eds Drauglis, E. & Jaffee, R. I.) 53–71 (Plenum, New York, 1975).
- Andrew, S. P. S. *Chem. Engng Sci.* **36**, 1431–1445 (1981).
- Asscher, M. & Somorjai, G. A. *Surf. Sci.* **143**, L389–L392 (1984).
- Broughton, J. Q. & Gilmer, G. H. *J. chem. Phys.* **79**, 5119–5127 (1983).
- Frenken, J. W. M. & van den Veen, J. F. *Phys. Rev. Lett.* **54**, 134–137 (1985).
- Schrammen, P. & Holzl, J. *Surf. Sci.* **130**, 203–228 (1983).
- Huttig, G. F. *Arch. Metallk.* **2**, 93–99 (1948).
- Nature **315**, 628 (1985).
- Rideal, E. K. & Taylor, H. S. *Catalysis in Theory and Practice*, 175–177 (Macmillan, London, 1926).
- Galwey, A. K., Gray, P., Griffiths, J. F. & Hasko, S. M. *Nature* **313**, 668–671 (1985).
- Feidenhansl, R. & Stensgaard, I. *Surf. Sci.* **133**, 453–468 (1983).
- Hupkens, T. M. & Fluit, J. M. *Surf. Sci.* **143**, 267–286 (1984).
- Neihus, H. *Surf. Sci.* **130**, 41–49 (1983).
- Onuferko, J. H., Woodruff, D. P. & Holland, B. W. *Surf. Sci.* **87**, 357–374 (1979).
- Jones, G. J. R., Onuferko, J. H., Woodruff, D. P. & Holland, B. W. *Surf. Sci.* **147**, 1–14 (1984).
- Shih, H. D., Jona, F., Jepsen, D. W. & Marcus, P. M. *Phys. Rev. Lett.* **46**, 731–734 (1981).
- Bonzso, F., Ertl, G., Grunze, M. & Weiss, M. *J. Catal.* **49**, 18–41 (1977).
- Wu, N. L. & Phillips, J. J. *phys. Chem.* **89**, 591–600 (1985).
- Bettler, P. C., Bennum, D. H. & Case, C. M. *Surf. Sci.* **44**, 360–376 (1974).
- Kuijpers, E. G. M., Kock, A. J. H. M., Werner, G. J. & Geus, J. W. *J. Catal.* **88**, 1–7 (1984).
- Klabunde, K. J. & Imizu, Y. *J. Am. chem. Soc.* **106**, 2721–2722 (1984).
- Chinchen, G. C. *et al. Preprints Am. chem. Soc. Div. Fuel Chem.* **29**, 178–188 (1984).
- Friedrich, J. B., Wainwright, M. S. & Young, J. D. *J. Catal.* **80**, 1–13; 14–24 (1983).
- Baglin, E. G., Atkinson, G. B. & Nicks, L. J. *Ind. Engng Chem. Prod. Res. Dev.* **20**, 87–90 (1981).
- Pritchard, J., Catterick, T. & Gupta, R. K. *Surf. Sci.* **53**, 1–20 (1975).
- Corma, A., Martin, M. A. & Perez-Pariente, J. *JCS Chem. Commun.*, 1512–1513 (1983).
- Corma, A., Martin, M. A. & Perez-Pariente, J. *Surf. Sci.* **136**, L31–L34 (1984).
- Drissen, J. M., Poels, E. K., Hindermann, J. P. & Ponc, V. *J. Catal.* **82**, 26–34 (1983).
- Ichikawa, S., Poppa, H. & Boudart, M. *J. Catal.* **91**, 1–10 (1985).
- Hicks, R. F. & Bell, A. T. *J. Catal.* **90**, 205–220 (1984).
- Gonzalez, L., Miranda, R. & Ferrer, S. *Surf. Sci.* **119**, 61–70 (1982).
- Ozaki, A. *Accs chem. Res.* **14**, 16–21 (1981).
- Bossi, A., Garbassi, F. & Petri, G. *JCS Faraday Trans. 1* **78**, 1029–1038 (1982).
- Urabe, K., Yoshioka, T. & Ozaki, A. *J. Catal.* **54**, 52–56 (1978).
- Burch, R. & Garba, L. C. *React. Kinet. Catal. Lett.* **16**, 315–320 (1981).
- Del Angel, G., Coq, B., Dutartre, R. & Figueras, F. *J. Catal.* **87**, 27–35 (1984).
- Cox, M. P., Ertl, G., Imbihl, R. & Rustig, J. *Surf. Sci.* **134**, L517–L523 (1983).
- Tauster, S. J., Fung, S. C. & Garten, R. L. *J. Am. chem. Soc.* **100**, 170–175 (1978).
- Spencer, M. S. *J. Catal.* **93**, 216–223 (1985).
- Vannice, M. A. & Chou, P. in *Proc. 8th int. Congr. Catal. Vol. 5*, 99–110 (Chemie, Weinheim, 1984).
- Anderson, J. B. F., Bracey, J. D., Burch, R. & Flambard, A. R. in *Proc. 8th int. Congr. Catal. Vol. 5*, 111–121 (Chemie, Weinheim, 1984).
- Gregor, M., Ihme, B., Kotter, M. & Rieckert, L. *Ber. Bunsenges phys. Chem.* **88**, 427–433 (1984).
- Thomas, C. L. *Catalytic Processes and Proven Catalysts* (Academic, New York, 1970).

A gene required for the specification of dorsal-ventral pattern in *Drosophila* appears to encode a serine protease

Robert DeLotto & Pierre Spierer

Département de Biologie moléculaire, Université de Genève, 30, quai Ernest Ansermet, CH-1211 Geneva, Switzerland

The maternal effect gene snake is required for the establishment of the dorsal-ventral axis during the embryonic development of Drosophila. The molecular cloning of the gene and analysis of a complementary DNA sequence suggest that the gene encodes a serine protease which is structurally similar to proteases involved in blood clotting, peptide processing, and complement fixation pathways.

ONE of the long standing goals of developmental biology is to understand the mechanisms underlying pattern formation. The identification of mutations affecting the embryonic development of *Drosophila melanogaster* has rendered aspects of some pattern forming processes amenable to molecular analysis¹. One such process is the establishment of the dorsal-ventral axis. The dorsal-ventral axis of the *Drosophila melanogaster* embryo is established very early in embryogenesis by the action of the products of a relatively small number of genes^{2,3}. Most of the genes involved in this process are maternal-effect genes, that is, genes for which the phenotype of the embryo depends upon the genotype of its mother. Maternal-effect genes code for products such as messenger RNAs or proteins which are stored in the oocyte during oogenesis and used at later times during embryogenesis.

The 'dorsal group' genes are a group of eleven maternal-effect genes involved in the establishment of dorsal-ventral pattern³. Members of the group, which includes *snake*, share a common recessive phenotype which resembles that initially described for mutations in the gene *dorsal* (*dl*). Embryos from *dl/dl* females develop in such a way that normal ventro-lateral pattern elements are displaced or absent and replaced by dorsal pattern elements. In extreme cases the embryo develops into a hollow tube of dorsal ectoderm. The fact that the recessive phenotype for mutations in most of these genes is the same has led to the suggestion³ that dorsal group gene products are involved in a common process of dorsal-ventral axis establishment.

The dorsal-ventral axis is not fixed at fertilization, but becomes so at about the time of cellular blastoderm formation; the development of an embryo from a female homozygote dorsal group mutant can often be altered by the injection of either cytoplasm or poly(A)⁺ RNA from wild type donor embryos³. For six out of ten of the dorsal group genes, ventro-lateral pattern elements can be restored in this way, although with different efficiencies. The *snake* locus responds most dramatically to cytoplasmic rescue and in 20% of cases an adult fly can be produced. The *snake* rescuing substance appears to be very active; as little as 0.5% by volume of a wild-type egg's cytoplasm can completely restore normal ventrolateral pattern elements in an embryo from a *snake* homozygote mother (C. Nüsslein-Volhard, personal communication).

Here we present evidence which suggests that the *snake* locus encodes a serine protease that has structural similarities to known serine proteases such as kallikreins, blood clotting factors and components of the complement system of higher vertebrates.

Cloning of the *snake* locus

Previous work⁴ has genetically mapped the *snake* locus to the right arm of the third chromosome at cytogenetic position 87D10-12 (ref. 4), within an existing 'chromosomal walk' of 320

kilobases (kb) spanning the *rosy* and *Ace* loci⁵. We used P-element mediated germ-line transformation⁶ to identify a fragment from this interval containing the *snake* locus. A number of DNA restriction fragments of the interval were subcloned into a P-element vector (Fig. 1). These constructs were injected and tested as shown in Fig. 2a. The experiment exploits the ability of a functional copy of the *snake* gene to overcome sterility of *snk/snk* female homozygotes. Of four constructs tested only one, C413.5X, restored fertility. This construct has a 13.5 kb *XhoI* fragment (see Fig. 1b) which runs from a point within the *rosy* structural gene (left) to a point within the *hsc2* heat shock cognate gene (right)⁷. We examined whether restoration of fertility was associated with transformation and integration of C413.5X into the germ line. First we compared genomic DNA from 'transformed' and 'non-transformed' flies by Southern blot (Fig. 2b). Second, we used the technique of *in situ* hybridization to polytene chromosomes. Figure 2c shows an *in situ* hybridization of radioactively labelled C413.5X DNA to a polytene chromosome squash of a transformed fly. Both experiments show that the rescue of *snake* phenotype requires integration of C413.5X DNA into the germ line. The precise location of *snake* can be determined by mapping rearrangement breakpoints falling within the C413.5X insert. Figure 3 is a schematic map of the *rosy-hsc2* intergenic region. The *snake* gene must lie entirely within a 4.5 kb portion of this intergenic region, based on deletion mapping. The *ry*⁵⁰⁶ deletion lacks part of the *rosy* structural gene and ~1.5 kb of DNA downstream of *rosy*⁸, but homozygous *ry*⁵⁰⁶ flies are fertile. Therefore, *ry*⁵⁰⁶ cannot delete essential *snake* sequences. *Df(3R)ry*³⁶, a 35 kb deletion which breaks within the 2.4 kb *EcoRI* fragment, removes essential *snake* sequences. This suggests that part of the *snake* gene lies within the 2.4 kb *EcoRI* fragment (Fig. 3, centre). Two other rearrangements, *Df(3R)l26d*, a large deletion which removes *hsc2* DNA sequences and an insertion of an *Is101* element in *ry*¹¹ chromosomes do not destroy essential *snake* sequences.

Northern blot analysis

Poly(A)⁺ RNA was prepared from staged embryos at 0-2, 2-4, 4-6 and 10-12 hour intervals, electrophoresed in denaturing gels, transferred to membranes and probed with a radioactively labelled 2.4 kb *EcoRI* DNA fragment (Fig. 4a). There is a single transcript of 1,650 nucleotides in the 0-2 and 2-4 hour lanes but not between 4-12 hours.

Isolation of cDNAs

A 0-4 hour *D. melanogaster* cDNA library in a λ gt10 bacteriophage vector⁹ was screened with the labelled 2.4 kb *EcoRI* DNA fragment. Five hybridizing clones were isolated; the largest (cD8) will be discussed here. The position of cD8 is shown in

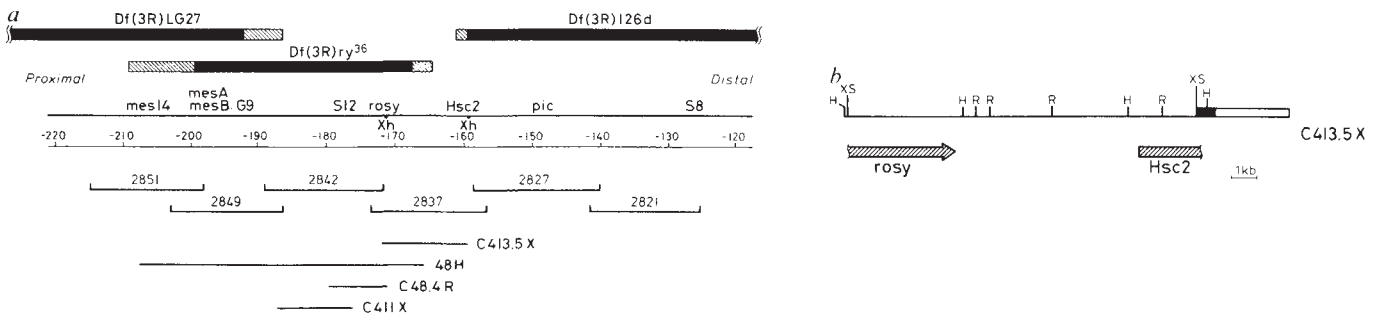


Fig. 1 *a*, A map of the 87D region including the *snake* locus; *b*, restriction map of recombinant C413.5X. In *a*, a number of chromosomal rearrangements have been mapped to DNA within the 87D region²⁰. *snake* is uncovered by the deletion *Df(3R)ry³⁶* but not by *Df(3R)LG27* nor by *Df(3R)126d*. This places the *snake* locus between the right hand breakpoint of *Df(3R)LG27* and the left hand breakpoint of *Df(3R)126d* with essential DNA within the deletion *Df(3R)ry³⁶*. Heavy bars (top), deletions which break within the region. Hatching, breakpoint limits (defined by restriction fragments). Complementation groups are indicated. The scale corresponds to the molecular walk coordinates from ref. 5. Overlapping recombinant bacteriophages are shown below. Bottom, inserts of four P-element constructs. Recombinants beginning with C are cloned into Carnegie-4 (ref. 6); 48H is a P-element cosmid clone isolated from a p-cos library²¹. In *b*, bar, insert DNA sequences; black, P-element sequences; white, vector. The limits of the *rosy* and *hsc2* structural genes are shown by hatched arrows. Scale bar, 1 kb.

Fig. 2 Germ line transformation of *snake*. *a*, selection scheme for transformants; *b*, Genomic Southern blot²³ of DNA from two different 'transformed' stocks. *c*, *In situ* hybridization of radioactively labelled C13.5X DNA to polytene chromosome squashes of *snk²²⁹* T13.5X/1. In *a*, Embryos from balanced stocks of the mutant *snk²²⁹ ru st e ca/TM3* (kindly provided by K. Anderson and C. Nüsslein-Volhard) were injected with each of the four constructs illustrated in figure 1 as described²². Females of the genotype *snk²²⁹ ru st e ca/snk²²⁹ ru st e ca* were isolated in subsequent generations and tested for fertility as indicated. Only females which integrate and express a wild type copy of the *snake* gene should be fertile. In *b*, Genomic DNA was isolated from 15 whole flies and digested with restriction endonuclease *Hind*III. Lane *a*, *snk²²⁹ ru st e ca/TM3* control flies; lane *b*, *snk²²⁹/snk²²⁹* T13.5X/1; lane *c*, *snk²²⁹/TM3* T13.5/1; lane *d*, *snk²²⁹/snk²²⁹* T13.5X/3; lane *e*, *snk²²⁹/TM3* T13.5X/3 probed with nick translated C13.5X DNA. The C13.5X construct contains *Hind*III sites at both sides of the insert within and near the P-element inverted repeat (see Fig. 1b). Chromosomes containing an additional copy of this DNA fragment due to transformation should exhibit two new fusion bands at 4.5 and 3.0 kb when probed with C13.5X as well as the three bands of 7.2, 6.4 and 2.95 kb from hybridization to 87D12. Furthermore, the same new pattern of hybridization is observed regardless of the sites of integration due to the unique position of *Hind*III sites adjacent to both points of integration. The two new bands appear in the 'transformed' flies at 4.5 and 3.0 kb (lanes *b-e*). The 3.0 kb band overlaps with the 2.95 kb band from 87D12. If the probe is washed off the filter and the filter is re-probed with the Carnegie-4 vector—which contains an a P-element from the white locus at 3C—(lanes *f-j*), a new band appears at 3.0 kb due to new hybridization to the P-element sequences introduced upon integration. *c*, A new site of hybridization is present in the transformed fly T13.5X/3 at the cytogenetic position 23AB.

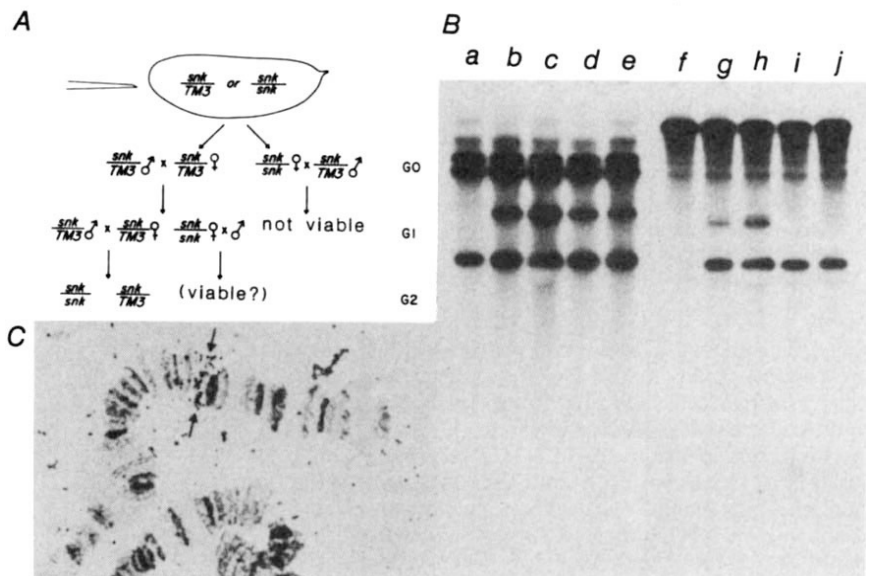
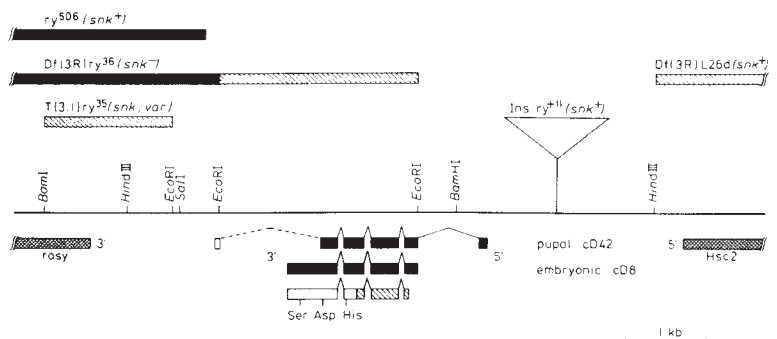


Fig. 3 Diagram of the *snake* locus. Dark bars (top), deleted DNA sequences in deficiencies. Hatched regions, restriction fragments in which the rearrangement breaks. Crosshatched bars (bottom), limits of the *rosy* and *hsc2* structural genes. The embryonic cD8 cDNA and a pupal cDNA are mapped relative to restriction sites indicated on the centre line. *T(3:1)ry³⁵* is a rearrangement which breaks within the 3' end of the *rosy* structural gene.



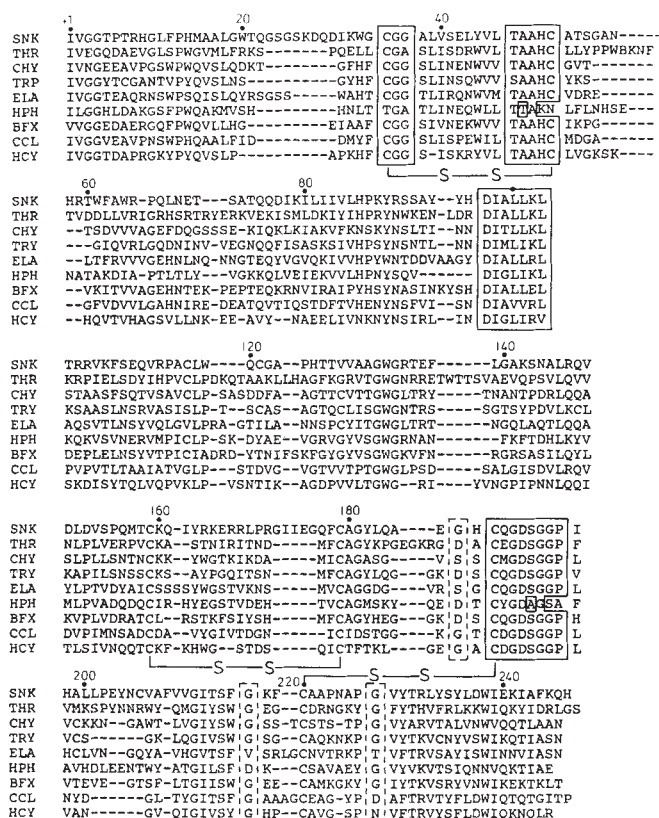


Fig. 6 An alignment of amino-acid residues in the carboxy-terminal catalytic regions of some serine proteases. SNK, part of the conceptual translation of cD8; THR, bovine thrombin; CHY, bovine chymotrypsin; TRP, bovine trypsin; ELA, porcine elastase; HPH, human haptoglobin heavy chain³⁰; BFX, bovine clotting factor X³¹; CCL, crab collagenase³²; HCY, hornet chymotrypsin³³. Boxes, conserved residues near the active sites. Dashed boxes, residues involved in substrate binding. HPH contains the sequence IRHYEGSTPEKKTPKSPVGGPILNEHTF inserted between positions 178 and 179. The standard one letter code for amino acids is used.

is a serine protease. Serine proteases are proteolytic enzymes that share a common catalytic mechanism and very well defined structural characteristics (for a general review see ref. 12). In particular, all members of this family share three strongly conserved amino acid sequence blocks that make up the active site of the enzyme. These sequences surround the catalytic His, Asp and Ser residues. In Fig. 6, the three reactive residues lie at positions 50, 97 and 193, respectively. Similarities near these active site residues are indicated by solid boxes. Human haptoglobin heavy chain (HPH) is the only member of the family that has no proteolytic activity¹³; it also has amino acid changes in critical active site residues. The catalytically active chains of serine proteases often begin with the amino acid sequence IVGG, marking the point at which the precursor is cleaved to produce an active protease. The sequence of the predicted protein contains IVGG at position 184. Six cysteine residues are conserved in all these proteins except HPH. These cysteines often form the three disulphide bridges shown in Fig. 6. Table 1 shows that the overall homology between SNK and these proteases is between 29% and 35%. This is about the same as the homology between any other two members of the group in this area. Since the active site residues, cysteine residues and overall hydrophilic and hydrophobic domains of SNK are all well conserved, we conclude that the protein encoded at the *snake* locus is a serine protease. However, a direct demonstration of enzymatic activity is required.

Table 1 Homology between catalytic regions of known serine proteases and the carboxy terminal amino acids of SNK

Protease	Identity	% homology
Bovine Thrombin	70/233	30
Bovine Chymotrypsin	71/220	32
Bovine Trypsin	71/217	33
Porcine Elastase	71/226	31
Human Haptoglobin	67/224	30
Bovine Factor IX	66/225	29
Crab Collagenase	69/218	32
Hornet Chymotrypsin	74/210	35

A number of possibilities arise if the *snake* product is a serine protease. Serine proteases are typically made as zymogens which are then activated by specific proteolytic cleavage. The amino-terminal residues are important in the activation and in some cases in the localization of the enzyme. These amino termini may be short, as for the digestive enzymes, or quite long with diverse functions, as for prothrombin. The amino-terminal portion of the *snake* protein (183 residues) is similar in length to that of blood clotting Factor IX (181 residues). Upon activation (cleavage) of many serine proteases the amino terminus remains linked to the catalytic chain by a disulphide bridge. SNK may have a similar disulphide bridge and subunit structure. The first 25 amino acids of the predicted protein are hydrophobic and may function as a signal peptide (as for blood clotting Factor IX). The Ca²⁺ binding domain homology at the amino terminus of SNK is in a similar position to the Ca²⁺ binding domains of some blood clotting factors. Thus, the overall structure of the protein resembles that of a serine protease much like some of the blood clotting factors.

The possibility that the *snake* product may be made as an inactive precursor has implications for the mechanism of the pattern forming process. The mother may be able to provide the oocyte with the *snake* product in a stable storage form. The initial events of embryogenesis are rapid and apparently do not require expression of the zygotic genome. The production of a zymogen would allow multiple points for post-translational regulation. One or more of the other dorsal group gene products might activate the zymogen.

The positions of the amino acids in serine proteases that line the binding pocket and determine the substrate specificity are known¹⁴. These amino acids are at positions 189, 216 and 226 in chymotrypsin¹⁵ and for SNK, positions 187, 217 and 227, respectively in Fig. 6. These positions are all occupied by glycine in SNK (Fig. 6, dotted boxes). This arrangement is unlike that of any known serine protease. Glycines at these positions should result in a wide binding pocket like that of chymotrypsin. As glycine is not charged it is probable that SNK will not cleave at charged residues, unlike trypsin, which prefers to cleave after two positively charged residues.

Anderson and Nüsslein-Volhard³ have proposed that the products of the dorsal group genes interact to form the dorsal-ventral pattern, and have further suggested a model in which the *Toll* gene product is produced as an inactive precursor, then activated by the product(s) of the other dorsal-group genes¹⁶. A structural similarity between the *snake* product and the blood clotting and complement fixation components might imply some similarity of mechanism. For example, some of the products of the dorsal group genes might function in proteolytic cascades analogous to those in blood clotting^{17,18} and complement fixation¹⁹ in higher vertebrates. This would allow the *snake* product to be involved in the activation of the *Toll* gene product and provide a mechanism for *snake* to interact with the other dorsal group gene products. Such a cascade could be of localized effect and coupled with diffusion provide a way to generate and 'fix' positional information. Clearly, further biochemical

characterization of the *snake* gene product as well as characterization of the other dorsal group gene products will be necessary.

We thank A. Tissieres in whose laboratory this work was done, C. Tonka-Chen for technical assistance with the sequencing, O. Jenni and Y. DeLotto for the figures, C. Nüsslein-Volhard and K. V. Anderson for stocks and communicating unpublished

results, M. Noll for the cDNA library, M. Goldschmidt-Clermont for helpful discussions, S. Mayfield for helpful comments and P. Castrischer for the croissants. R.D. was supported by the European Molecular Biology Organization and the Swiss National Science Foundation. This work was supported by the Swiss National Science Foundation.

Received 21 July; accepted 10 September 1986.

1. Nüsslein-Volhard, C. & Wieschaus, E. *Nature* **287**, 795-801 (1980).
2. Nüsslein-Volhard, C. in *Determinants of Spatial Organization* (eds Subtelney, S. & Konigsberg, I. R.) 105-211 (Academic, New York, 1979).
3. Anderson, K. V. & Nüsslein-Volhard, C. in *Pattern Formation* (eds G. M. Malacinski & S. Bryant) 269-289 (Macmillan, New York, 1984).
4. Anderson, K. V. & Nüsslein-Volhard, C. *Nature* **311**, 223-227 (1984).
5. Bender, W., Spierer, P. & Hogness, D. S. *J. molec. Biol.* **168**, 1-35 (1983).
6. Rubin, G. & Spradling, A. *Nucleic Acids. Res.* **11**, 6341-6351 (1983).
7. Craig, E., Ingolia, T., Slater, M., Manseu, L. Bardwell, J. in *Heat shock from Bacteria to Man* (eds Schlessinger, M. J., Ashburner, M. & Tissieres, A.) 11-18 (Cold Spring Harbor, N.Y., 1980).
8. Clark, S. H., McCarron, M., Love, C. & Chovnick, A. *Genetics* **112**, 755-767 (1986).
9. Frei, D. *et al. Cold Spring Harb. Symp. quant. Biol.* **50**, 127-134 (1986).
10. Shepherd, J. C. W. *Proc. natn. Acad. Sci. U.S.A.* **78**, 1596-1600 (1981).
11. Collins, J. H., Greaser, M. L., Porter, J. D. & Horn, M. J. *J. biol. Chem.* **252**, 6356-6362 (1977).
12. Stroud, R. M. *Sci. Am.* **231**, 74-88 (1974).
13. Kurosky, A. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 3388-3392 (1980).
14. Kraut, J. A. *Rev. Biochem.* **46**, 331-358 (1977).

15. Bode, W., Schwager, P. & Huber, R. *Miami Winter Symposium* **11**, 43 (1976).
16. Anderson, K. V., Bokla, L. & Nüsslein-Volhard, C. *Cell* **42**, 791-798 (1985).
17. Davie, E. W. & Fujikawa, K. A. *Rev. Biochem.* **44**, 799-829 (1975).
18. Jackson, C. M. & Nemerson, Y. A. *Rev. Biochem.* **49**, 765-811 (1980).
19. Ried, K. B. M. & Porter, R. R. A. *Rev. Biochem.* **50**, 433-464 (1981).
20. Spierer, P., Spierer, A., Bender, W. & Hogness, D. *J. molec. Biol.* **168**, 35-50 (1983).
21. Haenlin, M., Steller, H., Pirrotta, V. & Mohier, E. *Cell* **43**, 827-837 (1985).
22. Rubin, G. & Spradling, A. *Science* **218**, 348-353 (1982).
23. Southern, E. M. *J. molec. Biol.* **98**, 503-517 (1975).
24. Maniatis, T., Fritsch, E. F. & Sambrook, J. in *Molecular Cloning, A Laboratory Manual* 202-203 (Cold Spring Harbor Laboratory, 1982).
25. Fyrberg, E. A., Kindie, K. L., Davidson, N. & Sodja, A. *Cell* **19**, 365-368 (1980).
26. Zimmerman, J. L., Petri, W. & Meselson, M. *Cell* **32**, 1161-1170 (1983).
27. Sodja, A., Arking, R. & Zafar, R. *Dev. Biol.* **90**, 363-368 (1982).
28. Sanger, F. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5464 (1977).
29. Henikoff, S. *Gene* **28**, 351-359 (1984).
30. Greer, J. J. *J. molec. Biol.* **153**, 1027-1042 (1981).
31. Titani, K. *et al. Proc. natn. Acad. Sci. U.S.A.* **72**, 3082-3086 (1975).
32. Grant, G. A., Eisen, A. Z. & Bradshaw, R. A. *Meth. Enzym.* **80**, 722-734 (1981).
33. Jany, K. D. & Haug, H., *FEBS Lett.* **158** 98-102 (1983).

LETTERS TO NATURE

Is there a common origin for the cosmic γ -ray lines at 0.51 and 1.81 MeV near the galactic centre?

W. R. Webber, V. Schönfelder & R. Diehl

Max-Planck-Institut für extraterrestrische Physik,
D-8046 Garching, FRG

The discovery of the 1.81-MeV ^{26}Al decay line in the present interstellar medium¹ provided direct experimental proof that intermediate mass nuclei are being synthesized continuously in the Galaxy. The instruments making this and a subsequent observation² had broad angular responses, thus, although both observations were consistent with a source near the centre of the Galaxy, interpretations for the origin of this radiation have included objects distributed throughout the Galaxy capable of synthesizing ^{26}Al , such as type II supernova, Wolf Rayet stars, novae and red giants. New observations³ have localized this source to the galactic centre region consistent with a point source at the centre within the resolution function of the telescope ($\sim 10^\circ$ FWHM, full width at half maximum). Observations of the galactic 0.51-MeV positron annihilation line have also localized this line to the galactic centre with an angular resolution $\sim 15^\circ$ FWHM⁴. Time variability requires that it come from a region ~ 1 pc in size, hence it is probably produced near a compact object at or near the galactic centre⁵. Here we discuss the possibility that the 0.51- and 1.81-MeV lines have a common origin: the initial decay of ^{26}Al to ^{26}Mg with the most probable emission of a positron which eventually annihilates producing the 0.51-MeV line, followed by the decay of ^{26}Mg to the ground state with the emission of the 1.81-MeV line. For this process to be solely responsible for both lines the time-integrated intensities of the lines must be compatible, and we show that this is not contradicted by the existing data. The measurement of the properties of the two lines individually and taken together place strong constraints on the source region of this radiation.

The radioactive isotope ^{26}Al has a half-life of 7.4×10^5 yr, decaying into an excited state of ^{26}Mg 82% of the time by emission of a positron and 18% of the time by electron capture. ^{26}Mg decays to its ground state 99.7% of the time by emission

of a 1.81-MeV γ -ray. There is, therefore, a relationship between the time-integrated intensity of the 0.51-MeV positron annihilation line and the 1.81-MeV line depending on how the positrons annihilate. Studies of the continuum emission observed at energies near 0.51 MeV suggest that ~ 50 - 90% of the positron annihilations take place through positronium^{4,6,7}, although the fraction is uncertain because of the difficulty in evaluating the underlying continuum. Even a fraction of 0% cannot definitely be excluded from the existing data. In the case of positronium annihilation, the three photon decay of the triplet state produces a continuum, whereas the singlet state annihilates into two 0.51-MeV photons thus giving a total of 0.5 photons at 0.51 MeV and 2.25 photons in the three photon continuum per initial positron. In the case of direct annihilation, there would be two 0.51-MeV photons per positron. If 50-90% of the positrons decay by positronium annihilation one would expect a yield of between 0.65 and 1.25 photons per positron. This factor times the ratio 0.82/1.00 should be the time-integrated ratio of the 0.51- to 1.81-MeV line intensities yielding a total of 0.5 to 1.0 photons per positron, respectively.

The time history of the two lines is shown in Fig. 1. There is no evidence of a time variability of the 1.81-MeV line over a period of ~ 3 yr, as might be expected, so we assume that this line has had a constant intensity $\sim 6 \times 10^{-4} \gamma \text{ cm}^{-2} \text{ s}^{-1}$. For the 0.51-MeV line the details of the line variability are not known—the variation could be periodic or aperiodic. Therefore, we consider only the period of the best and most complete measurements after late 1977. For ~ 2 yr the positron line intensity was $1.5 \times 10^{-3} \text{ cm}^{-2} \text{ s}^{-1}$. At that point in 1980 the line intensity dropped below the sensitivity limit of $4 \times 10^{-4} \gamma \text{ cm}^{-2} \text{ s}^{-1}$ and remained so at least through the end of 1984. If we assume that during this time the intensity approached zero, then the time-integrated 0.51-MeV intensity from late 1977 to late 1984 is $\sim 5 \times 10^{-4} \gamma \text{ cm}^{-2} \text{ s}^{-1}$, or about the same as the 1.81-MeV line. If the two lines arise from the same process then the 0.51-MeV line intensity may remain low for several more years at least, and the observed episode from 1977 to 1979 could have represented a period when many of the accumulated positrons from ^{26}Al decay found the right conditions to annihilate, the decrease in 1980 occurring possibly because the reservoir of positrons was used up.

Let us now examine how the properties of the two lines can

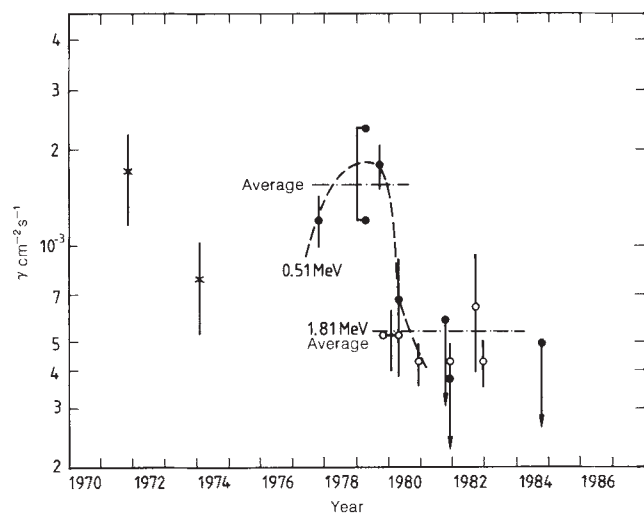


Fig. 1 Time history of 0.51- and 1.81-MeV line intensities. The 0.51-MeV data are summarized in refs 19, 20. The 1.81-MeV data are from refs 1–3. Note that the intensity quoted in ref. 1 is $\gamma \text{ cm}^{-2} \text{ s}^{-1} \text{ rad}^{-1}$. Crosses, 0.51 MeV, low resolution; filled circles, 0.51 MeV, high resolution; open circles, 1.81 MeV.

help to elucidate their possible common source region. In the case of the 0.51-MeV line, its narrow line width of $<2.5 \text{ keV}$ requires that it decay in a region where the bulk velocity along the line of sight must be $<200 \text{ km s}^{-1}$ and the gravitational redshift $z < 7 \times 10^{-4}$ (ref. 5). The line width $<3.5 \text{ keV}$ for the 1.81-MeV line leads to similar limits on the bulk velocity and the gravitational redshift¹ and thus is consistent with a common origin region for the two lines.

The observed intensity variation of the 0.51-MeV line coupled with its narrow width and line to continuum ratio constrains the size of the annihilation region, the temperature, density and possibly ionization fraction⁸ of the matter in it. The size of the region should obviously not exceed $\sim 1 \text{ pc}$ and the density and temperature should be high enough so that positrons can slow down and annihilate in less than a year. This size constraint excludes most of the multiple extended sources such as supernova and nova. In evaluating the time constants involved in positron annihilation we consider that the positrons have thermalized. The decay of thermal positrons is governed by the following types of processes: (1) direct annihilation with free electrons or neutral hydrogen; or (2) the formation of positronium with subsequent annihilation either through charge exchange with neutral hydrogen or through radiative recombination with free electrons. At temperatures $<8,000 \text{ K}$ practically no free electrons are available, therefore direct annihilation with neutral hydrogen with a relatively low annihilation cross-section is dominant. Above 10^4 K the charge-exchange process dominates, resulting in an annihilation rate that increases more than three orders of magnitude reaching a maximum at $T = 1.5 \times 10^4 \text{ K}$ for a gas in ionization equilibrium⁵. A natural mechanism to produce a positron episode would be to produce a temperature increase in a small cool and fairly dense region containing ambient thermalized positrons. If these positrons were in a medium with $n \sim 10^3 \text{ cm}^{-3}$ in ionization equilibrium, for example, the annihilation time would be only $\sim 10^7 \text{ s}$ for $T = 1.5 \times 10^4 \text{ K}$, $5,000\times$ shorter than for a unionized gas of the same density at a temperature $<8,000 \text{ K}$.

Even if the assumption of thermal equilibrium was invalid, our above argument concerning the time constants still applies. This can be seen in a unionized medium where the slowing down time from 1 MeV to $<100 \text{ eV}$ (ref. 9) at which the positrons will annihilate in flight, turns out to be of the same order as the annihilation time of thermal positrons in the same medium.

For the 1.81-MeV line, the fact that it does not seem to be accompanied by other observable lines from radioactive elements with shorter half-lives (for example, ^{22}Na at 1.275 MeV and ^{56}Ni at 0.81 MeV)¹ suggests that the ^{26}Al was not synthesized recently but could have been synthesized up to $\sim 10^6 \text{ yr}$ ago. The ^{26}Al line intensity requires that $5 \pm 2 \times M_{\odot}$ of ^{26}Al was synthesized within this timescale of $\sim 10^6 \text{ yr}$ (ref. 3). If the source or sources of this ^{26}Al synthesize it with: (1) the same mass fraction (2.6×10^{-4}) as calculated for nova ejecta¹⁰; (2) the mass fraction as calculated for Wolf-Rayet type ejecta (6×10^{-6})¹¹; or (3) the mass fraction as calculated for supernova like ejecta ($\sim 10^{-6}$)¹², then anywhere from 10^5 to a few times $10^6 M_{\odot}$ of newly synthesized material must also be present nearby.

To put this situation in perspective we briefly discuss conditions near the centre of our Galaxy. Infrared observations indicate a virial mass $4 \times 10^6 M_{\odot}$ within $\sim 1 \text{ pc}$ of the centre and a stellar mass increase αr at least out to $\sim 5 \text{ pc}$ (refs 13, 14). Thus $\sim 10^7 M_{\odot}$ is concentrated within a few parsecs of the centre. The gas within $1\text{--}2 \text{ pc}$ appears to be almost completely ionized, but outside this distance, the gas in the plane is neutral¹⁵. Observations at 2 and 6 cm show a spiral like pattern in the central $1\text{--}2 \text{ pc}$. The average density (n_e) in this pattern is $\sim 4 \times 10^3 \text{ cm}^{-3}$ (ref. 16). Only at the very centre do velocities exceed 200 km s^{-1} . As many as 14 small, compact ionized clouds with $n_e \sim 10^5 \text{ cm}^{-3}$, $d \sim 0.2 \text{ pc}$ and $T_{\text{eff}} \sim 3 \times 10^4 \text{ K}$ have been observed in the central few parsecs¹³.

A common source region explaining the observed properties of both lines can thus be found in the region within a few parsecs from the centre (excluding perhaps the central parsec or less) where bulk velocities, densities, ionization conditions, temperatures and so on seem to be within the range required to explain the properties of both the lines. The following scenario illustrates the relationships between the parameters involved for the production of the two lines. Suppose an impulsive source of energy (such as photons) heats and ionizes a small region with $r \sim 0.2 \text{ pc}$ within a larger region where the ^{26}Al and positrons are uniformly distributed. The positron annihilation rate in this region may increase by a factor up to 5×10^3 to its maximum at $T = 1.5 \times 10^4 \text{ K}$ where $T_{\text{ann}} n = 5 \times 10^9 \text{ s atom cm}^{-3}$ (ref. 5). At this time, the peak intensity of the 0.51-MeV positron annihilation line is observed which is ~ 2.5 times the steady 1.81-MeV ^{26}Al line intensity (for example, 1977–79). Before this episode, the accumulated positrons will build up to a steady-state density which depends on the quiescent annihilation rate, however, the total rate of quiescent positron decay, which is $\sim n^+ / T_{\text{ann}}$, will lead to an annihilation line production rate which is $\sim 0.5\text{--}1.0$ times the ^{26}Al decay rate as noted earlier. During the episode the positrons in this small region are rapidly used up and the 0.51-MeV line disappears. To produce the observed increase of the 0.51-MeV line of 2.5 times the steady 1.8-MeV line it is required that the ratio of this small region to the overall volume containing the ^{26}Al must be simply $\Delta V/V \sim (2.5 - 1) r_V / r_{\Delta V} \sim 3 \times 10^{-4}$, assuming one 0.51-MeV photon is produced per positron (here $r_{\Delta V}$ and r_V refer to the annihilation rates in both the small and large regions, respectively). Thus, the radius of a spherical region containing the ^{26}Al would be $\sim 0.2 \times (3 \times 10^3)^{1/3}$ or $\sim 3 \text{ pc}$. This is obviously a very simplified relative size scale illustration. If the conditions for positron annihilation are less favourable, for example, ionization equilibrium is not reached in the timescales involved, then the volume ratio will tend to be larger. On the other hand, the larger region containing the ^{26}Al is probably not homogeneous but has a higher density in the localized regions containing the positron episodes, thus balancing the above effect. Overall one might envision a larger volume within which several 0.51-MeV line episodes are possible during an extended period of time as other small regions have the local positron annihilation rate impulsively modified. Some of these episodes may be strong and outshine the quiescent 511-keV flux (like the one which caused the high flux between 1977 and 1979), others may be weaker

and hardly detectable.

In this model, all positrons required to explain the present 511-keV flux are provided by the decay of ^{26}Al . Because the positrons have—in quiet conditions—a mean annihilation time of $\sim 1,000$ yr, the event that produced the ^{26}Al must have occurred at least 1,000 yr ago otherwise positrons from other radioactive isotopes produced in the same event would also contribute. Regardless of the details of specific models for the production of ^{26}Al or positrons, there must be a relationship between the quiescent 0.51-MeV line intensity, and the peak 0.51-MeV line intensity relative to that from ^{26}Al , and the relative volumes of the regions involved in the respective line emissions that are a function of the increase in positron annihilation rate. Because of the size limitations imposed by these relationships the ^{26}Al cannot be produced in explosive events that distribute it too widely (~ 10 – 100 pc) in $\sim 10^6$ yr near the centre of the Galaxy.

On the other hand, the region containing the ^{26}Al cannot be too small, otherwise the abundance ratio of ^{26}Al (n_{Al}) relative to H (n_{H}) would become unacceptably high. For example, if $4 M_{\odot}$ of ^{26}Al is confined within a volume with $r = 3$ pc, then $n_{\text{Al}} \sim 5 \times 10^{-2} \text{ cm}^{-3}$. If n_{H} within this volume is 10^3 cm^{-3} then $n_{\text{Al}}/n_{\text{H}} \sim 5 \times 10^{-5}$. Though this ratio is already rather high, it is of the same order as that found in galactic cosmic-ray sources¹⁷.

A common origin for the two lines is possible, provided the ^{26}Al is really concentrated within a localized region near the centre as is suggested by the recent measurements³. A detailed astrophysical model based on this idea will be published elsewhere. However, whatever the source, or sources, of the 0.51-MeV and 1.81-MeV lines, active nucleosynthesis is occurring near the centre of our Galaxy—perhaps in types of objects not

yet fully considered by theorists. To try and understand better the possible relationships between these two important lines it is necessary to make simultaneous measurements of both lines at a sensitivity level $\sim 10^{-4} \gamma \text{ cm}^{-2} \text{ s}^{-1}$, so that the 'quiescent' 0.51-MeV line from positron decay, which must surely be present at some level between 0.5 and 1.0 times the ^{26}Al line intensity, can be observed. Such measurements can be made, also with improved angular resolution, with COMPTEL¹⁸ on the γ -ray observatory (GRO) of NASA.

W.R.W. thanks the Alexander von Humboldt-Stiftung for a Senior US Scientist Award. We thank Greg Morfill for useful discussions.

Received 27 May; accepted 15 August 1986.

1. Mahoney, W. A., Ling, J. C., Wheaton, W. A. & Jacobson, A. S. *Astrophys. J.* **286**, 578–585 (1984).
2. Share, G. H. *et al. Astrophys. J.* **292**, L61–L65 (1985).
3. von Ballmoos, P., Diehl, R. & Schönfelder, V. *Astrophys. J.* (submitted).
4. Leventhal, M., MacCallum, C. J. & Stang, P. D. *Astrophys. J.* **L14** (1978).
5. Lingenfelter, R. E. & Ramaty, R. *The Galactic Center, AIP Conf. Proc.* **83**, 148–159 (1982).
6. Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. A. & Jacobson, A. S. *Positron-Electron Pairs in Astrophysics, AIP Conf. Proc.* **101**, 230–235 (1983).
7. Riegler, G. R., Ling, J. C., Mahoney, W. A., Wheaton, W. M. A. & Jacobson, A. S. *Astrophys. J.* **294**, L13–L15 (1985); erratum **305**, L33 (1986).
8. Brown, B. L. *Astrophys. J.* **292**, L67–L70 (1985).
9. Bussard, R. W., Ramaty, R. & Drachman, R. J. *Astrophys. J.* **228**, 928–934 (1979).
10. Hillebrandt, W. & Thielemann, F.-K. *Astrophys. J.* **255**, 617–623 (1982).
11. Dearborn, D. S. P. & Blake, J. B. *Astrophys. J.* **288**, L21–L24 (1985).
12. Clayton, D. D. *Astrophys. J.* **280**, 144–149 (1984).
13. Lacy, J. H. *The Galactic Center, AIP Conf. Proc.* **83**, 53–59 (1982).
14. Genzel, R. *et al. Astrophys. J.* **276**, 551–559 (1984).
15. Mezger, P. G. & Wink, J. E. *Astr. Astrophys.* **157**, 252–266 (1986).
16. Oort, J. H. *The Milky Way Galaxy, IAU Symp.* **106**, 349–366 (1985).
17. Webber, W. R. *Astrophys. J.* **252**, 386–392 (1982).
18. Schönfelder, V. *et al. IEEE Trans. nucl. Sci.* **31**, 766–770 (1984).
19. Leventhal, M., McCallum, C. J., Hutters, A. F. & Stang, P. D. *Astrophys. J.* **260**, L1–L5 (1982).
20. MacCallum, C. J. & Leventhal, M. *Proc. 19th Int. Cosmic Ray Conf.* **1**, 213–216 (1985).

Sputtering of sodium on the planet Mercury

M. A. McGrath*, R. E. Johnson* and L. J. Lanzerotti†

* University of Virginia, Charlottesville, Virginia, USA

† AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

The discovery of a column density of neutral sodium vapour ($\sim 10^{11}$ – 10^{12} atoms Na cm^{-2}) associated with the solar-facing side of the planet Mercury¹ has rekindled interest in the possible presence of atmospheric gases on that planet. Recent observations² confirm the presence of Na emission, with its centroid shifted to the sunward limb. The sodium fluorescence is observed in the visible range. Previous ultraviolet observations by Mariner 10 provided evidence for the presence of H and He in the atmosphere ($\sim 10^{11}$ and 10^{12} atoms cm^{-2} respectively³) and gave upper limits on other possible species (such as O, Ne, Ar and C). Whereas the H and He can be explained primarily by capture from the solar wind⁴, we show here that this is not so for Na by evaluating the depletion rates. Photon and ion bombardment can release Na from its normal bonding state, producing adsorbed and ejected Na in ballistic trajectories. We show that ion sputtering cannot account for the observed column density but is an important loss mechanism for Na. Photons are likely to be the dominant stimulus, both directly through photodesorption and indirectly through thermal desorption of adsorbed Na. We therefore conclude that the atmosphere produced is characterized by the planet's surface temperature, with the ion-sputtered Na contributing to a lesser, but more extended, component of the atmosphere.

We take the exospheric column density of the observed Na to be $N \approx \Phi_s \tau$ where Φ_s is the supply flux of Na to the 'atmosphere' and τ is the lifetime of a Na atom in this atmosphere. If the source is a sputter (desorption) source we can write

$\Phi_s = \Phi_i f \bar{Y}_i$ where Φ_i is the stimulating radiation flux of type i , $\bar{Y}_i$ is the surface-averaged yield and f is the fractional atomic concentration of Na in the surface layer. Table 1 contains our estimates for f , Φ_i , $\bar{Y}_i$ and τ , and the resulting column densities for ion and photon sputtering of Na. The calculated column densities are in general smaller than required by the observations, with photon sputtering providing the largest atmosphere. **Surface composition (f).** Although the surface composition of Mercury is not well known, analogies are most often made with lunar materials. Based on a typical composition for lunar rocks⁵ we estimate the atomic fraction of Na on the surface to be $f \approx 0.002$, which implies a surface significantly depleted in alkalis. Sodium is likely to be bound to oxygen (for example, in $\text{NaAl}_2\text{Si}_2\text{O}_8$ feldspar) with a binding energy $U \approx 2$ eV. If Mercury remained active after formation, this would tend to enhance the concentration of Na-containing minerals in the crust. In addition, meteorite action in the upper metres of the surface can refresh the surface through overturning of the regolith. In the upper centimetres irradiation by energetic particles can enhance the diffusion of Na in and through the surface layer down to the penetration depths of the most energetic ions⁶. For example, under high electron fluxes diffusion of Na is observed in glasses, a likely component of the surface⁷. Under low-flux, but long-term, ion bombardment the more volatile species would tend to diffuse preferentially to the surface, enhancing the Na concentration, some of which may exist as adsorbed Na. Finally, exogenic sources such as the solar wind and meteoritic material can contribute Na.

Ejected Na atoms recondensing on the surface would generally be adsorbed with reduced bonding to the surface because the initially attached O would make new and/or tighter bonds with the other surface species. On a rough surface of rocky material containing Na the ejection and readorption of the species is enhanced twofold over other, less volatile, species⁸. In the presence of a small O atmosphere, on the other hand, the adsorbed Na and O atoms could combine chemically at the

Table 1 Sputtered atmospheric column densities

	Collisional sputtering		Electronic sputtering		
	H ⁺	He ²⁺	H ⁺	He ²⁺	$h\nu$
$\bar{Y}^*$	~ 0.1	~ 1	≤ 0.04	≤ 0.02	$\leq 10^{10}-10^{11}$
Φ (cm ⁻² s ⁻¹)	$\sim 10^8-10^9$	$\sim 10^7-10^8$	$\sim 10^8-10^9$	$\sim 10^7-10^8$	
N (atoms Na cm ⁻²)†	$2 \times 10^7-2 \times 10^8$	$2 \times 10^7-2 \times 10^8$	$\leq 10^7-10^8$	$\leq 10^6-10^7$	

* $U = 2$ eV was used as a rough estimate of the chemical binding energy. The estimates ignore the possibility of absorbed Na and enhancements due to higher surface temperatures.

† N obtained from Eqs (1) and (2) with $f \approx 0.002$ and $\tau \approx 10^3$.

surface. There are therefore a number of processes that affect the state of the Na. In the absence of more detailed knowledge of the surface, we used $f \approx 0.002$ and $U \approx 2$ eV to describe the ejection of Na.

Sputtering (desorption) yield (Y). When either electronic or collisional energy is deposited by incident radiation, atoms or molecules can be ejected from the surface. We divide the processes in Table 1 into two categories according to the type of energy deposition. The yield depends on the environment of the species of interest (that is, surface composition, roughness (or porosity)⁹ and temperature), the energy density deposited near the surface, and the nature of the binding of the species to the surface. At low temperatures and low excitation densities, Y becomes independent of temperature and linear in the deposited energy density. The yield increases with increasing angle of incidence to the normal, which we ignore to approximately account for roughness. For this minimum-yield case $\bar{Y}_i = \bar{\Delta z}_i / \lambda_i$, where $\bar{\Delta z}_i$ is a weighted sputter depth and λ_i is the mean path between energizing events¹⁰.

For collisional sputtering by ions, $\bar{\Delta z}_i$ is proportional to U^{-1} and λ_i^{-1} is proportional to S_n , the nuclear stopping (energy loss) cross-section. For a material consisting of atoms of similar masses, $Y \approx c\alpha S_n / U$, where $c = 4.2 \times 10^{14}$ cm⁻² and α is a factor depending primarily on the mass of the incident particle¹¹. We use $\alpha = 1$ and $S_n = 5 \times 10^{-16}$ and 3×10^{-15} eV cm² for H and He respectively. This sputtering process is characterized by an energy spectrum of the ejected Na of the form $E/(E+U)^3$, which determines the ballistic flight times and the gravitational escape fraction.

Sputtering due to the electronic energy deposited by ions, electrons and photons occurs when this energy produces, directly or indirectly, repulsive states between a surface atom and a neighbour. Here $\lambda_i^{-1} \approx n\sigma_i$, the atomic number density times the excitation cross-section. As ejection occurs from the first monolayer, $\bar{\Delta z}_i \approx P_e n^{-1/3}$. Because electronic recombination of a localized ionization in insulators occurs repulsively, σ_i is roughly the 'ionization' cross-section for the binding electrons and $P_e \leq 0.1$ (refs 10, 12–15). For electron-stimulated desorption (~ 50 -eV electrons, $\sigma_i \approx 10^{-16}$ cm²), using $n^{2/3} \approx 2 \times 10^{15}$ atoms cm⁻² gives $Y \leq 0.02$, consistent with the results for desorption of Na from alkali halides¹³. For energetic ions $\sigma_i \approx S_e / W$, where S_e is the electronic stopping cross-section ($S_e \approx 3 \times 10^{-15}$ and 2×10^{-14} eV cm² for H and He, respectively) and W (~ 20 – 30 eV) is the mean energy required to produce an electron-hole pair at an oxygen site. For solar photon-stimulated desorption we can use the photodissociation rate (J) of, for instance, NaO₂ to estimate $\bar{Y}\Phi_i$. At Mercury, $\Phi_i \sigma_i = J \approx 10^{-3}$ – 10^{-4} s⁻¹ (ref. 16), which is used with P_e above. We also combine the appropriate solar photon flux with the extrapolated photodesorption yields from the work by Tolk and Larchuk¹³ on NaCl, to find an estimate of $\bar{Y}\Phi_i$ consistent with the lower value of J above.

A warm surface, such as the sunlit hemisphere of Mercury, can increase the electronic desorption yield beyond the values given here¹⁷. For temperature T , the yield can be written as $Y[1 + A \exp(\Delta E/kT)]$, where ΔE is an activation energy (~ 0.2 – 0.4 eV for liberated Na from alkali halides^{13,14}). In effect,

the penetrating ions and photons dissociate species within the solid and cause the segregation of the more volatile species at the surface. This is known to occur for many alkali-containing species^{13,14,17,18}. For the case we are considering, the Na and O (it would be N in an azide, halide in an alkali halide) would be energetically separated by the process described above and the Na would be desorbed later. That is, subsequent bombardment enhances the transport of Na to the surface and the ejected atoms would have an energy distribution which is primarily characterized by the surface temperature with a small energetic contribution determined by direct repulsive ejection. This is also consistent with the observations for a porous surface^{8,9}.

Lifetime (τ). In a ballistic atmosphere ($\approx 3 \times 10^{15}$ atoms cm⁻²) above a surface, the lifetime of an ejected neutral species is determined by its average time of flight or the ionization rate. The average solar photoionization lifetime at Mercury is $\tau_i \approx 8,000$ s (refs 19, 20). We assume 2 kT as an average desorption energy for electronic sputtering, giving $\tau_s \approx 500$ s. For collisionally ejected atoms of mass 23, use of the average energy $\frac{1}{2}U$ gives $\tau_s \approx 1,000$ – $2,000$ s. Solar radiation pressure on the Na can reduce these times considerably, depending on the distance from the Sun^{20,22}. On the other hand, the sputtered species will probably not fully accommodate to the warm surface on their return²³. Studies by Hapke and co-workers^{8,9} indicate that the relative sticking coefficient for Na onto silicate (the major constituent of the lunar regolith) is very small (~ 0.05 – 0.24). The lifetime of a particle may therefore be an order of magnitude larger than the ballistic lifetime, in which case ionization limits the lifetime of Na. In the following, we use a conservative average τ of order 10^3 s.

Particle radiation (Φ). Ions and electrons can potentially impact the surface of Mercury in several ways. We discuss the two most important processes here.

(i) Nominal solar-wind particles ($n \approx 50$ cm⁻³, 95% 1 keV H⁺ and 5% 7 keV He²⁺) can be expected to impact directly on the surface approximately 6% of the time⁴. The largest contribution hitting the planet, however, comes from the precipitation of solar-wind particles through the expected cusps in the magnetosphere and along 'closed' field lines, to which the particles have access by diffusion and/or by reconnection processes at the magnetopause. Based on estimates given by Goldstein *et al.*⁴ we take the total particle flux incident on the surface from these processes as $\approx 1.5 \times 10^8$ cm⁻² s⁻¹.

(ii) Energetic particles produced in substorm-like processes in the magnetosphere, primarily in the planet's magnetotail, can be directed inward and guided along magnetic field lines toward the polar regions. These particles can also be diffused radially across night-side field lines rather easily by the existing plasma-wave turbulence until they reach the point where their finite gyroradii cause a direct impact on the planet. This process is more difficult to estimate because relatively little is known about Hermean substorm processes other than the measurements of energetic electrons (and possibly ions) made by Mariner 10 (ref. 24). The Helios-2 spacecraft reportedly made measurements of energetic protons (~ 87 – 176 keV) that escape from the Hermean magnetosphere²⁵. These particles were attributed to substorm

activity and/or escape from the 'radiation belt'. Extrapolated to the planet's bow shock, the flux of these particles was $\sim 3 \times 10^4 \text{ cm}^{-2} \text{ s}^{-1} \text{ ster}^{-1}$ or $3 \times 10^5 \text{ cm}^{-2} \text{ s}^{-1}$ omnidirectional. If we assume the front-side bow shock to be the surface of a sphere with radius $\sim 3R_M$ (the dawn- and dusk-side distances are larger than the subsolar distance) and that these particles cover a hemisphere, the fluxes can be extrapolated to the planet's dimension. We can reasonably assume that an equal number of substorm, magnetotail-produced particles can impact the night side of the planet; the fluxes would then be $\sim 3 \times 10^6 \text{ cm}^{-2} \text{ s}^{-1}$. If the spectrum of accelerated ions varies as E^{-2} to E^{-3} , then the flux of $E \approx 10 \text{ keV}$ protons available to impact the surface would be of order $3 \times 10^8 \text{ cm}^{-2} \text{ s}^{-1}$, or possibly as high as $10^9 \text{ cm}^{-2} \text{ s}^{-1}$. The total proton flux used is then approximately 10^8 – $10^9 \text{ cm}^{-2} \text{ s}^{-1}$. More energetic particles from an active Sun can penetrate the magnetic barrier efficiently. However, we estimate their net effect to be insignificant.

As can be seen from Table 1, unless the quantities estimated above are changed drastically, ion sputtering, originally suggested by Potter and Morgan, cannot account for the observed column density. The UV photo-desorption process, which is the least reliable estimate at present, gives the largest column density. Longer-wavelength photons can also contribute to this column density by thermal stimulation of Na. This is equivalent to assuming that the accommodation to the surface is not unity. If the Na is indeed chemisorbed to the surface (rather than chemically bonded to O) then thermal desorption is likely to be an efficient means of producing the observed column density. Therefore photons, via photo desorption and thermal desorption, are likely to account for the observed column density. However, the impacting ions also play an important role because they enhance the diffusion of Na to the surface and separate Na from O. We note that in this process a related amount of O will also be released. As returning O atoms will more efficiently accommodate to the surface, a column density of O $\leq 10^{11} \text{ cm}^{-2}$ is expected, within the limits set by Mariner 10 (ref. 26). We therefore suggest that the Na atmosphere is characterized by the surface temperature of Mercury, with a smaller and much more extended component produced by ion sputtering.

We can now estimate the loss of Na from Mercury. The escape fraction from the atmosphere due to electronic sputtering is negligible. Based on the energy distribution given above, this fraction is ~ 0.7 for collisional sputtering. This produces a loss rate $\sim 10^5$ – $10^6 \text{ Na cm}^{-2} \text{ s}^{-1}$, which is fairly independent of the binding assumed. Atmospheric Na is also lost by ionization. That is, the Na^+ can reach open magnetic field lines and be lost from the atmosphere and magnetosphere⁴. Here we write the escape flux as $\beta N/\tau_i$ where β is the fraction of Na^+ that can access open field lines and τ_i is the lifetime of the Na against ionization as given above. For He, $\beta \approx 0.5^4$. The larger mass and the solar radiation pressure imply lower ballistic trajectories for Na. We therefore use $\beta \leq 0.1$. From the observed column density, the depletion due to ionization processes ($\sim 10^6 \text{ Na cm}^{-2} \text{ s}^{-1}$) and collisional sputtering (above) are roughly comparable. Consequently, both the character of the magnetic field at Mercury and the ballistic transport of Na are important for the Na loss.

Sodium has not been specifically identified in the solar wind. If its abundance is similar to the identified elements such as Mg, Ne or Si^{27–29}, then its total fractional content of the solar wind would be $\sim 10^{-4}$. This supply rate ($< 10^5 \text{ Na cm}^{-2} \text{ s}^{-1}$) is much less than the total loss rate of Na from Mercury. Therefore, unlike the He atmosphere, the presence of a Na atmosphere implies that Mercury is being depleted in Na at a rate equivalent to a loss in 10^9 yr of all of the Na in the top 1–10 m, based on $f \approx 0.002$. This would suggest a transport mechanism in the upper layers of the planet, such as recycling by micrometeorites, to maintain a Na concentration at the surface, and/or an exogenic source such as micrometeoritic material. Although ion bombardment does not produce the observed atmosphere it is

likely to be important in determining the depletion rate of Na and hence in helping to constrain models for Mercury's surface.

We acknowledge helpful conversations with W. Smyth, N. Tolk, K. B. McAfee, J. Bell, T. E. Madey and M. Nelson. R.E.J. acknowledges the partial support of AT&T Bell Laboratories and the Planetary Geosciences Division of the University of Hawaii. The work at the University of Virginia is supported by the NASA Solar System Exploration Division under grant NAGW-186 and the NSF Astronomy Division under grant AST-85-11391.

Received 25 February; accepted 11 August 1986.

- Potter, A. & Morgan, T. *Science* **229**, 651–653 (1985).
- Schneider, N. M., Wells, W. K. & Hunten, D. M. *Bull. Am. astr. Soc.* **17**, 712 (1985).
- Broadfoot, A. L., Kumar, S., Belton, M. J. & McElroy, M. B. *Science* **185**, 166–169 (1974).
- Goldstein, B. E., Seuss, S. T. & Walker, R. J. *J. geophys. Res.* **86**, 5485–5499 (1981).
- Vinogradov, A. P. in *Cosmochemistry of the Moon and Planets*, NASA spec. Publ. No. 370 (PI), 5–33 (1977).
- Murti, D. K. & Kelly, R. *Thin solid Films* **33**, 149–163 (1976).
- Miotello, A. & Mazoldi, P. *Phys. Rev. Lett.* **54**, 1675–1678 (1985).
- Cassidy, W. & Hapke, B. *Icarus* **25**, 371–383 (1975).
- Hapke, B. *Icarus* **66**, 270–279 (1986).
- Johnson, R. E. & Brown, W. L. *Nucl. Instr. Meth.* **198**, 103–118 (1982).
- Sigmund, P. in *Sputtering by Ion Bombardment* Vol. 1 (ed. Berisch, E.) 9–71 (Springer, New York, 1981).
- Johnson, R. E., Lanzarotti, L. J. & Brown, W. L. *Adv. Space Res.* **4**, 9, 41–51 (1984).
- Haglund, R. F. *et al. Nucl. Instr. Meth.* (in the press).
- Tolk, N. & Larchuk, T. S. *Inst. Phys. Conf. Ser.* no. 71, 135–141 (1984).
- Leung, C., Vass, M. & Gomer, R. *Surf. Sci.* **66**, 67–71 (1977).
- Kirchoff, V. W. *Geophys. Res. Lett.* **10**, 721–724 (1983).
- Kelly, R. *Surf. Sci.* **90**, 380–318 (1979).
- Evans, B. L., Yoeff, A. D. & Gray, P. *Chem. Rev.* **59**, 515–568 (1959).
- Swider, W. *Planet. Space Sci.* **17**, 1233–1246 (1969).
- Carlson, R. W., Matson, D. L. & Johnson, T. V. *Geophys. Res. Lett.* **2**, 469–471 (1985).
- Smyth, W. H. *Bull. Am. astr. Soc.* **17**, 712 (1985); *Nature* **323**, 696–699 (1986).
- Ip, W.-H. *Geophys. Res. Lett.* **13**, 423–426 (1986).
- Shemansky, D. D. & Broadfoot, A. L. *Rev. Geophys. Space Phys.* **15**, 491–499 (1977).
- Simpson, J. A., Eraker, J. H., Lampton, J. E. & Walpole, P. H. *Science* **85**, 160–166 (1974).
- Kirsch, E. & Richter, A. K. *Ann. Geophys.* **3**, 13–18 (1985).
- Kumar, S. *Icarus* **28**, 579–591 (1976).
- Geiss, J. in *Proc. ESA Workshop on Future Missions in Solar, Heliospheric, and Space Plasma Physics*, ESA SP-235, 37–50 (1985).
- Kunz, S., Bochsler, P., Geiss, J., Ogilvie, K. W. & Caplan, M. A. *Sol. Phys.* **88**, 359–376 (1983).
- Bame, S. J., Asbridge, J. R., Feldman, W. C., Montgomery, M. D. & Kearney, P. D. *Sol. Phys.* **43**, 463–473 (1975).

Nature and variability of Mercury's sodium atmosphere

W. H. Smyth

Atmospheric and Environmental Research, Inc., 840 Memorial Drive, Cambridge, Massachusetts 02139-3758, USA

Our knowledge of Mercury has been derived almost exclusively from the analysis of a limited amount of data acquired by the Mariner 10 spacecraft during its three flybys of the planet (29 March 1974, 21 September 1974 and 16 March 1975). The surprising discovery of a sodium atmosphere on Mercury from ground-based observations¹ has not only re-kindled interest in the planet but has also opened, through continuing ground-based observations^{2,3}, a new and promising way to explore the planet, its atmosphere and its interactions with the solar wind. Here I examine the peculiar behaviour and likely spatial character of the sodium atmosphere. Because of the large and extremely variable solar radiation pressure experienced by sodium atoms and the suggested presence³ of a non-thermal surface source, the size and shape of the sodium atmosphere can vary tremendously with Mercury's orbital location about the Sun, thus providing valuable signatures for remote study.

The atmosphere of Mercury, as deduced from Mariner 10 data, is very tenuous, with an upper limit⁴ for the surface gas density on the day side of $\sim 10^6 \text{ cm}^{-3}$. With such a low surface density, Mercury's atmosphere is an exosphere, as gas atoms and molecules travel in primarily ballistic trajectories and very rarely collide, except with the planetary surface. Of the atmos-

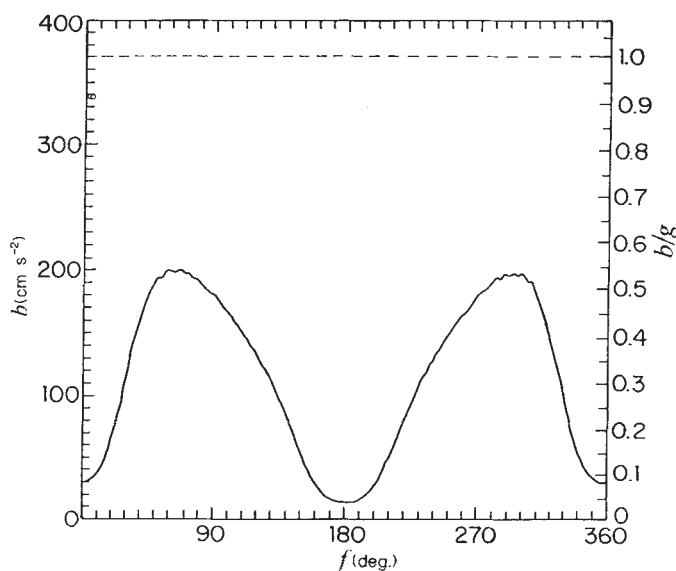


Fig. 1 Solar radiation pressure of sodium atoms near Mercury. The radiation acceleration (b) experienced by sodium atoms in Mercury's atmosphere because of resonance scattering in the solar D-lines is shown in absolute and surface gravity units as a function of the true anomaly angle f for the planet on its elliptical orbit around the Sun.

pheric constituents, only helium and hydrogen were positively identified, and atomic oxygen very tentatively identified, with estimated⁴ subsolar point densities of $4,500 \text{ cm}^{-3}$, 8 cm^{-3} and $7,000 \text{ cm}^{-3}$, respectively. The helium and hydrogen extended several thousands of kilometres above the surface and reflected at these altitudes surface velocity distributions which, although different, were roughly thermal (that is, approximately characterized by the surface temperature of 575 K). A non-thermal (cooler) component of hydrogen was, however, also observed near the surface, providing a total apparent surface hydrogen number density of 90 cm^{-3} . The abundances of He and H are roughly understood in terms of an approximate balance between sources and loss processes. The dominant sources appear to be outgassing of He from the planetary interior⁵ and the eventual capture and neutralization of solar wind ions (He^{++} , H^+) by the planetary surface, which is usually⁵⁻⁸ largely shielded from direct interactions with the variable conditions of the solar wind by a planetary magnetic field which forms a magnetopause at a nominal subsolar altitude of $\sim 1,460 \text{ km}$ (0.6 Mercury radii)⁹. The dominant sink is likely to be photoionization of neutrals followed by their loss (when not recaptured by the surface) through magnetospheric convection^{5,10}. Although the He and H atmospheric data have received a moderate amount of study¹⁰⁻¹³, a major uncertainty inhibiting future progress is the lack of understanding of the non-maxwellian (and hence non-barometric) character of the atmosphere produced by the gas-surface interaction¹⁰⁻¹² because Mercury's atmosphere is an exosphere.

Sodium gas in the sunlit atmosphere of Mercury, initially detected¹ through ground-based high-resolution spectrographic observations of its optical emissions in the D_2 (5889.95 Å) and D_1 (5895.92 Å) lines excited by solar resonance scattering, has since been independently verified³ and observed on several additional occasions (ref. 2 and A. E. Potter and N. M. Schneider, personal communications). The column abundance of sodium in the atmosphere, originally estimated¹ to be $\sim 8 \times 10^{11} \text{ atoms cm}^{-2}$, was revised downwards² (A. E. Potter, personal communication) to $(1-2) \times 10^{11} \text{ atoms cm}^{-2}$, implying a sodium optical depth in the D-lines of order unity. If the sodium atoms at the surface have an approximately thermal velocity distribution (characterized by the surface temperature of 575 K, or

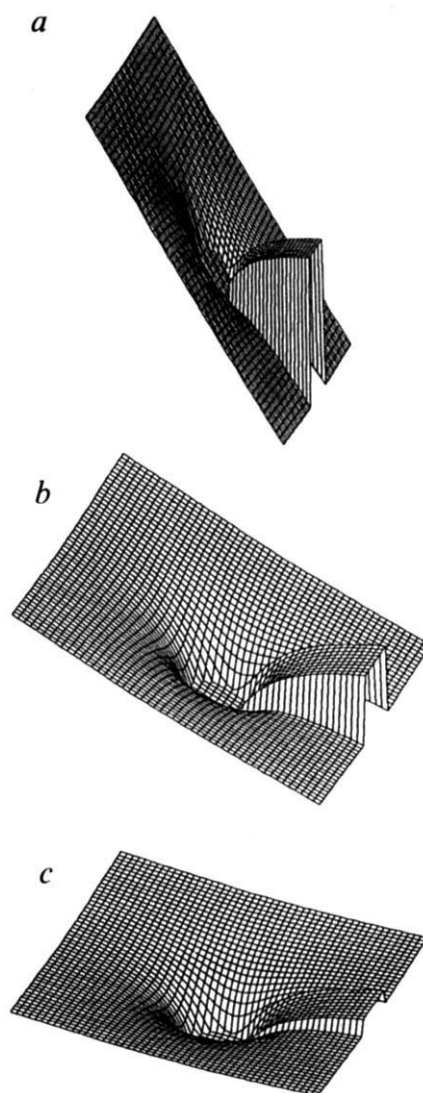


Fig. 2 Effective potential for a sodium atom near Mercury. The effective potential Φ may be expressed in terms of a dimensionless potential $\Phi^* = -1/r^* - b/g x^*$ upon division by GM/R , where G is the universal gravitational constant, M is the planetary mass and R is the planetary radius. Here $r^* = (x^{*2} + y^{*2} + z^{*2})^{1/2}$ and x^* , y^* and z^* are spatial coordinates, centred on the planet and measured in units of R , with the x^* -axis directed along the Sun-Mercury line and with the subsolar point at $x^* = -1$. The first term of Φ^* describes the radially dependent and directed gravitational acceleration of Mercury. The second term of Φ^* describes the constant and Sun-planet-directed solar radiation acceleration, where $g (=371 \text{ cm s}^{-2})$ is the surface gravity and b is the solar radiation acceleration shown in Fig. 1. The potential Φ^* is evaluated about the planet on a plane containing the x^* -axis and is shown in a surface representation for three different locations of Mercury on its orbit and the Sun: a , $f = 64^\circ$; b , $f = 150^\circ$; and c , $f = 180^\circ$. The viewing perspective and vertical scale for the three potential surfaces are the same. The Sun is to the left and the great circle where the plane intersects the planet is the bottom perimeter of the potential well. The impact of the planetary shadow on Φ^* is seen to the right of the potential well, where $b = 0$.

equivalently by a mean atomic speed of 0.6 km s^{-1}), the sodium atmosphere would have an exospheric surface number density of $\sim (2-4) \times 10^4 \text{ cm}^{-3}$ and be confined to within $\sim 100 \text{ km}$ of the surface. Some observations³, however, have suggested that sodium extends $\sim 700 \text{ km}$ above the surface at the subsolar point, implying the existence of non-thermal atomic velocities of at least 2 km s^{-1} and a broader velocity distribution, with speed components probably exceeding the surface escape speed of

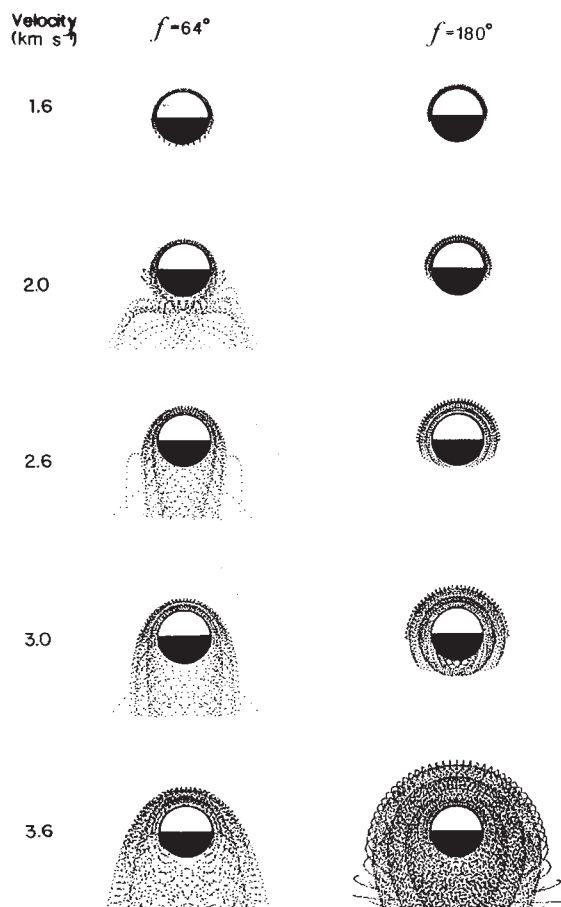


Fig. 3 Spatial character of the sodium atmosphere. The shape and size of the atmosphere for five sodium ejection speeds and for Mercury at true anomaly angles of 64° and 180° are shown in a dot-plot representation. A dot occurs at a time interval of 225 s along each of the 259 atom trajectories used in the atmospheric simulation. For the field of view selected, flight times of most non-planet-colliding trajectories were 1–3 h, with a maximum value of 4 h imposed.

4.3 km s^{-1} . Such a non-thermal source would obviously provide a new way to investigate the nature of the gas-surface interaction. Both magnetospheric-ion sputtering and photo-sputtering of sodium from the surface have been suggested as possible source mechanisms^{1,14–16}, with the former providing a non-thermal component¹⁵.

The lifetime of sodium in Mercury's atmosphere is very short and is dominated by photoionization. Because Mercury's orbital path about the Sun is an ellipse with a significant eccentricity (0.2056), the photoionization lifetime¹⁷ of sodium varies from $\sim 1.4 \text{ h}$ at perihelion (0.3075 AU) to $\sim 3.3 \text{ h}$ at aphelion (0.4667 AU). If sodium were lost from the atmosphere at this rate, the observed column abundance¹ of $\sim (1\text{--}2) \times 10^{11} \text{ atoms cm}^{-2}$ would imply a surface loss rate of $\sim (1\text{--}2) \times 10^7 \text{ atoms cm}^{-2} \text{ s}^{-1}$ or a sunlit hemispheric source rate of $\sim (4\text{--}8) \times 10^{24} \text{ atoms s}^{-1}$. The actual surface loss rate is likely to be significantly less, however, because, on ionization, many sodium atoms will not be lost by magnetospheric convection but will be recycled by impact with the planetary surface^{5,15}. With 90% recycling, for example, the loss rate could be balanced by meteoritic vaporization¹⁸ of surface material if the vapour contained $\sim 1\%$ sodium. The rate at which sodium ions are created above the surface and that at which sodium atoms are ejected by ion-sputtering from the surface could be coupled, as atoms with altitudes below the magnetopause at the subsolar point

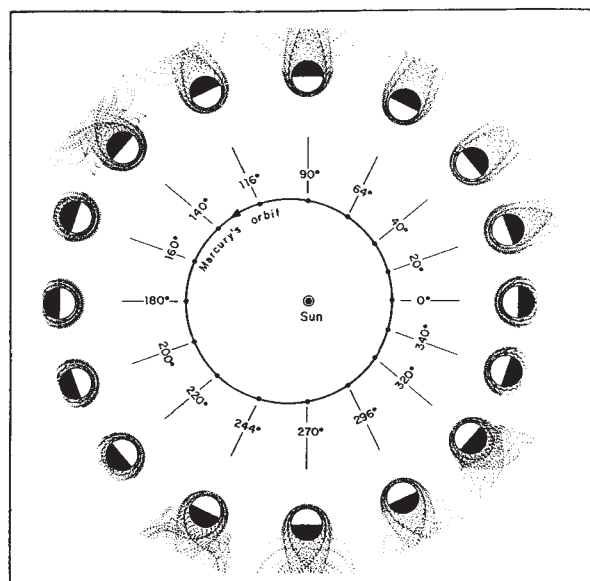


Fig. 4 Variation of the atmosphere with Mercury's orbital position. The shape and size of the sodium atmosphere for an ejection velocity of 2.6 km s^{-1} are shown for Mercury at 16 symmetric and diametrically opposite locations around the Sun. The dot-plot representation is described in Fig. 3.

have ballistic flight times (a fraction of an hour) and photoionization times which are comparable.

The motion of sodium atoms in Mercury's atmosphere is much more complex than that of He, H or O atoms, because the solar radiation acceleration experienced by sodium in the D-line is large^{16,19} and extremely variable¹⁹. The variability results from the fact that the amount of sunlight available to sodium varies with both the changing distance and radial velocity (due to solar absorption features) of Mercury on its elliptical orbit around the Sun. The acceleration, shown in Fig. 1 as a function of the true anomaly angle f of Mercury, varies from a maximum value of $\sim 200 \text{ cm s}^{-2}$ (a surprising 54% of surface gravity) near $f = 60^\circ$ and $f = 300^\circ$ to a minimum value of $\sim 12 \text{ cm s}^{-2}$ (3.3% of surface gravity) at aphelion ($f = 180^\circ$). Because the lifetime of sodium is extremely short (1.4–3.3 h) compared with the orbital period of Mercury ($\sim 88 \text{ days}$), the dynamical behaviour of sodium atoms in the atmosphere will adjust essentially instantaneously to the changing solar conditions at the planet. The morphology of the sodium atmosphere can thus be expected to vary significantly and exhibit an $\sim 88\text{-day}$ period.

The dramatic effects of solar radiation pressure on sodium in Mercury's atmosphere are displayed by examining the effective potential Φ for atoms that have zero velocity relative to the planet (the classical regions of motion analysis²⁰). Figure 2 shows the two-dimensional nature of this potential, including the planetary shadow, for the maximum ($f = 64^\circ$), an intermediate ($f = 150^\circ$), and the minimum ($f = 180^\circ$) value of the solar radiation acceleration. For an atom moving with a velocity of a few kilometres per second relative to the planet, the force experienced is similar to and only slightly different from that determined from Φ , as will be evident in the complete velocity-dependent treatment (including the planetary shadow) adopted below.

Figure 3 is an accurate illustration of the shape of Mercury's sodium atmosphere, as determined by a well-known method^{21,22} of following the orbits of many atoms, which are here initially ejected isotropically from the sunlit surface in the planetary equator plane for five different velocities that range between the mean surface thermal speed of 0.6 km s^{-1} and the surface escape

speed of 4.3 km s^{-1} . For the minimum value of the solar radiation acceleration ($f=180^\circ$), the atmospheric shape departs significantly only for ejection velocities $\geq 3.0 \text{ km s}^{-1}$ from what it would be if the solar radiation acceleration (b) were zero. For the maximum value of the solar radiation pressure ($f=64^\circ$), the atmosphere is severely depressed at the subsolar point for all ejection velocities because of the larger 'effective acceleration' of gravity ($b+g$), and for velocities as low as 2.0 km s^{-1} , atoms escape the atmosphere and form a long comet-like neutral sodium tail which is largely protected from photoionization by the planetary shadow. The spatial character of the atmosphere for the complete range of the true anomaly f is shown in Fig. 4 for a nominal ejection speed of 2.6 km s^{-1} . In Fig. 4, the subsolar height of the atmosphere ranges between a possible altitude³ of $\sim 700 \text{ km}$ ($f=296^\circ$) and the magnetopause altitude of $\sim 1,460 \text{ km}$ ($f=180^\circ$). A comparison of the atmospheric shape for symmetrically opposite locations above and below the (horizontal) aphelion-perihelion line indicates a real orbital asymmetry which occurs because the solar radiation acceleration is dependent on the atomic velocity.

The spatial character and density of the sodium atmosphere can therefore be expected to be extremely variable and periodic (~ 88 days). Additional periodic behaviour may also occur because of the 2:3 commensurability between the axial and orbital periods of Mercury, if the source of sodium on the surface is non-uniform. Future ground-based observations providing

improved spatial information for the sodium atmosphere will allow us to estimate more correctly the flux and velocity distribution of atoms at the surface and to explore more fully the nature of all atmospheric constituents and the interaction of the planet with the solar wind.

Received 24 March; accepted 24 July 1986.

1. Potter, A. E. & Morgan, T. H. *Science* **229**, 651-653 (1985).
2. Potter, A. E. & Morgan, T. H. *Bull. Am. astr. Soc.* **17**, 711-712 (1985).
3. Schneider, N. M., Wells, W. K. & Hunten, D. M. *Bull. Am. astr. Soc.* **17**, 712 (1985).
4. Broadfoot, A. L., Shemansky, D. E. & Kumar, S. *Geophys. Res. Lett.* **3**, 577-580 (1976).
5. Goldstein, B. E., Suess, S. T. & Walker, R. J. *J. geophys. Res.* **86**, 5485-5499 (1981).
6. Suess, S. T. & Goldstein, B. E. *J. geophys. Res.* **84**, 3306-3312 (1979).
7. Slavin, J. A. & Holzer, R. E. *J. geophys. Res.* **84**, 2076-2082 (1979).
8. Hood, L. L. & Shubert, G. J. *J. geophys. Res.* **84**, 2641-2647 (1979).
9. Ness, N. F., Behannon, K. W., Lepping, R. P., Whang, Y. C. & Schatten, K. H. *Science* **185**, 151-160 (1974).
10. Shemansky, D. E. *J. geophys. Res.* **85**, 221-222 (1980).
11. Shemansky, D. E. & Broadfoot, A. L. *Rev. Geophys. Space Phys.* **15**, 491-499 (1977).
12. Smith, G. R., Shemansky, D. E., Broadfoot, A. L. & Wallace, L. J. *Geophys. Res.* **83**, 3783-3790 (1978).
13. Hodges, R. R. Jr *J. geophys. Res.* **85**, 164 (1980).
14. McGrath-Kinnally, M. & Johnson, R. E. *Bull. Am. Astr. Soc.* **17**, 712 (1985).
15. McGrath, M., Johnson, R. E. & Lanzerotti, L. J. *Nature* (in the press).
16. Ip, W.-H. *Geophys. Res. Lett.* **13**, 423-426 (1986).
17. Carlson, R. W., Matson, D. L. & Johnson, T. V. *Geophys. Res. Lett.* **2**, 469-472 (1975).
18. Grün, E., Zook, H. A., Fechtig, H. & Giese, R. H. *Icarus* **62**, 244-272 (1985).
19. Smyth, W. H. *Bull. Am. astr. Soc.* **17**, 712 (1985).
20. Szebehegy, V. *Theory of Orbits: The Restricted Problem of Three Bodies* (Academic, New York, 1967).
21. Smyth, W. H. & McElroy, M. B. *Astrophys. J.* **226**, 336-346 (1978).
22. Smyth, W. H. *Astrophys. J.* **264**, 708-725 (1983).

The electronic structure of the benzene molecule

David L. Cooper*, Joseph Gerratt† & Mario Raimondi‡

* Department of Inorganic, Physical and Industrial Chemistry, University of Liverpool, PO Box 147, Liverpool L69 3BX, UK

† Department of Theoretical Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK

‡ Dipartimento di Chimica Fisica ed Elettrochimica, Università di Milano, Via Golgi 19, 20133 Milano, Italy

The Kekulé description of benzene as a mixture of the two structures



was given a firm foundation in quantum theory as a 'resonance hybrid' (see, for example, ref. 1). As molecular orbital (MO) theory developed, it was felt that the aromatic character of benzene was explained more naturally in terms of delocalized orbitals. The view that delocalized electrons provide essentially the correct description for this type of system appears to be accepted at all levels²; however, we present here theoretical evidence which challenges this view. We show that the π -electrons in benzene are almost certainly localized and that the characteristic properties of such a system arise from the mode of spin coupling.

We have developed the spin-coupled valence bond description of molecular electronic structure³⁻⁵ (also ref. 6 and refs therein), in which a system of N electrons is described by N distinct non-orthogonal orbitals whose spins are coupled to the required overall resultant S according to a particular coupling scheme k . The final spin-coupled wave function is a linear combination of all the possible coupling schemes k . The orbitals are expanded in a basis set, much as in MO theory, and all the variational parameters are optimized simultaneously to minimize the energy.

In the MO description of benzene, the orbitals are orthogonal and the lowest energy state of the molecule is obtained when the individual MOs are doubly occupied. In general, the orbitals

Table 1 Total energies for benzene using $[3s\ 2p/2s]$ and $[3s\ 3p/2s]$ contracted gaussian basis sets

Calculation	Energy (hartree*)	
	$[3s\ 2p/2s]$	$[3s\ 3p/2s]$
SCF		
(Molecular orbital)	-229.99539	-230.00150
Spin-coupled		
(One structure)	-230.07043	-230.07627
Spin-coupled plus		
All ionic structures	-230.07739	-230.08324

* 1 hartree = $4\pi^2 me^4/h^2(4\pi\epsilon_0)^2 = 4.3598 \times 10^{-18} \text{ J}$.

form bases for irreducible representations of the point group of the molecule (D_{6h} in this case) and are necessarily delocalized. It is possible in certain cases, such as that of H_2O , to localize the MOs without changing the total wave function, although there is no unique method for doing so. The spin-coupled wave function is not constrained in any of these ways and is much more flexible. The non-orthogonality of the orbitals allows for characteristics entirely different from those of molecular orbital theory; for example, in the spin-coupled description of H_2O the orbitals are highly localized and we can identify individual O-H bonds. This localization arises as a result of electron correlation, and cannot be removed without completely altering the spin-coupled wave function, which is not invariant under a linear transformation of the orbitals.

Additional correlation may be included in the spin-coupled method by generating excited spin-coupled configurations and by diagonalizing the resulting hamiltonian matrix over these structures. The final spin-coupled valence bond (VB) wave functions are very compact and afford a great deal of physical and chemical visuality. Nonetheless, total energies and molecular properties are of comparable accuracy to those from the most sophisticated configuration interaction methods.

To describe the electronic structure of benzene, we first performed a standard molecular orbital self-consistent field (SCF) calculation using a basis set of double- ζ quality— $(10s\ 5p/4s)$ gaussian primitives contracted to $[3s\ 2p/2s]$ ⁷. The spin-coupled calculation was then performed with the 36 electrons of the σ

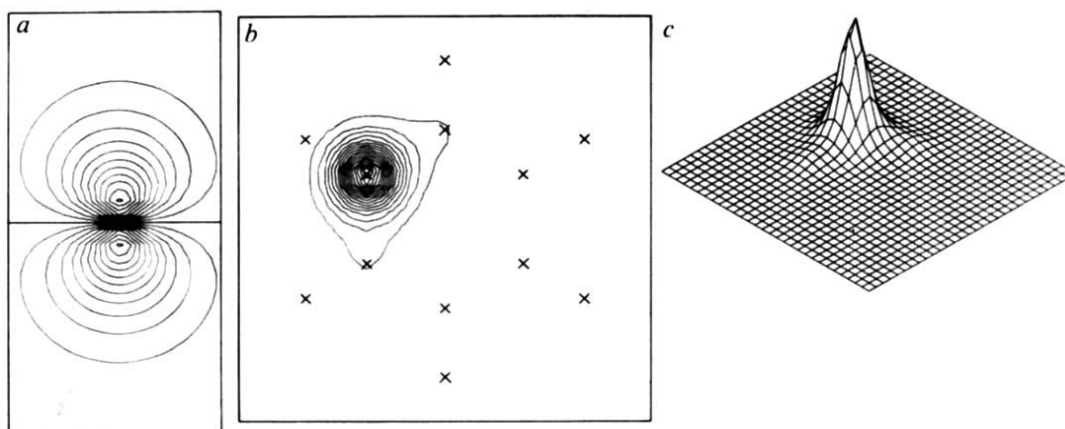


Fig. 1 One of the spin-coupled orbitals, ϕ_i , in benzene. The others are generated by successive C_6 rotations. *a*, Contours of $\phi_i(r)$ in a σ_v mirror (perpendicular to the molecular ring and containing two C-H bonds). *b*, Contours of $|\phi_i(r)|^2$ in a plane 1 Bohr radius (0.053 nm) above the C_6H_6 plane (the nuclei below are denoted by crosses). *c*, Isometric view corresponding to *b*.

framework accommodated in the doubly-occupied molecular orbitals. The 6 π -electrons were described by 6 non-orthogonal orbitals expanded in a basis set consisting of the 12 occupied and unoccupied π molecular orbitals. An advantage of this approach, in which we start from the delocalized SCF wave function, is that the results will most clearly highlight the differences between the spin-coupled and MO descriptions. The total wave function is of the form

$$\Psi = \sum_{k=1}^5 c_{0k} \mathcal{A} \{ \sigma_1^2 \sigma_2^2 \dots \sigma_{18}^2 \pi_1 \pi_2 \pi_3 \pi_4 \pi_5 \pi_6 \Theta_{0,0,f}^{36} \Theta_{0,0,k}^6 \} \quad (1)$$

in which the c_{0k} are the spin-coupling coefficients, $\mathcal{A}$ is the anti-symmetrizer, $\Theta_{0,0,f}^{36}$ is the spin function for the σ -electrons and $\Theta_{0,0,k}^6$ is a spin function for the π -electrons. Note that the spin functions, and thus the total wave function, are eigenfunctions of the operators for the square of the total spin $\hat{S}^2$, and of the projection of the spin onto an arbitrary, space-fixed z -axis, $\hat{S}_z$. For a system of six electrons with zero net spin, there are five possible spin couplings, all of which were included in the calculation. The π spin-coupled orbitals and the spin-coupling coefficients were optimized simultaneously, without any restrictions apart from the normalization of the orbitals.

The converged spin-coupled wave function consists of six highly localized orbitals, the spins of which are coupled symmetrically around the carbon ring framework to give the correct overall X^1A_{1g} state. All five spin functions contribute. The orbitals all possess an identical energy and shape, and they can be transformed into one another by successive C_6 rotations. As shown in Fig. 1, the orbitals are essentially deformed $C(2p_z)$ orbitals which are distorted symmetrically towards the neighbouring carbon atoms on each side.

As shown in Table 1, the total energy decrease relative to the m.o. solution is 0.0750 hartree. This is considerable. The energy optimization in the spin-coupled method is achieved by means of a second-derivative procedure known as the 'stabilized Newton-Raphson' method, with a maximum tolerance of 10^{-10} on any variational parameter; the energy is converged to many more figures than are quoted in Table 1. We have only very rarely experienced convergence to a subsidiary (local) minimum with this method, and it has always been obvious when this was the case, as the convergence characteristics are very different. This consideration, together with the size of the energy decrease relative to the SCF solution, makes it virtually certain that our present solution for benzene corresponds to the global minimum. Also shown in Table 1 are the results obtained with a slightly larger $[3s\ 3p/2s]$ basis set. In this case, the π -orbitals are expanded in 18 delocalized π MOs; the energy lowerings and the converged spin-coupled orbitals are very similar.

It is worth emphasizing the number of different solutions encompassed by wave function (1). One solution which could result is a modification of the MO description $a_{2u}^* e_{1g}^4$ with some

additional 'radial' correlation, that is, $\{a_{2u} a'_{2u} (e_{1g}^{(1)} e_{1g}^{(2)}) \times (e_{1g}^{(1)} e_{1g}^{(2)})\}$. In this picture, all the π -orbitals would be delocalized. Another possible solution included in wave function (1) is the alternant molecular orbital (AMO) function in which the six π -orbitals are divided into two sets of three with symmetries a_2'' and e'' of the subgroup D_{3h} of the molecular point group; the sets are interchanged by the operation C_6 (see refs 3, 8). Only one spin function may occur ($k=1$ in the Kotani scheme¹²), corresponding to the maximum spin ($S=3/2$) in each set and subsequent coupling of these to a net singlet. Many other solutions are possible, with varying degrees of delocalization.

The actual converged spin-coupled wave function corresponds to none of the cases just described, but to identical, highly localized orbitals on each of the six carbon atoms with the spins coupled symmetrically around the ring. On further reflection it is clear that a more delocalized solution for the orbitals necessarily entails a restriction on the number of participating spin functions. Evidence has also been presented, using m.o. theory, that delocalization of the π -electrons in benzene is energetically unfavourable⁹.

It must be emphasized that the SCF molecular orbitals of benzene are delocalized only because they are all constrained to have D_{6h} symmetry. This is a direct result of the restrictions of orthogonality and double occupancy. Indeed, in an SCF calculation with only a single $2p_z$ basis function on each carbon atom, the (delocalized) orbitals are completely determined by symmetry alone. The current calculations have been performed in a basis set of delocalized MOs and could easily have converged onto a solution with delocalized orbitals. However, when the restrictions of MO theory are lifted, the orbitals become highly localized.

Great use is made of the fact that SCF molecular orbitals belong to a given irreducible representation of the molecular point group. For example, this feature is very convenient for the interpretation of photoelectron spectroscopy (PES) experiments. However, there is no inconsistency between the spin-

Table 2 Occupation numbers of the spin functions in the Rumer basis

Rumer function	Occupation no.
1	0.4028
2	0.0648
3	0.0648
4	0.4028
5	0.0648

Functions 1 and 4 correspond to Kekulé structures, functions 2, 3 and 5 to Dewar structures.

coupled description of benzene and its PES spectrum. Removal of an electron from localized orbital φ_i produces a function Ψ_i which lacks the necessary spatial symmetry; one simply takes an appropriate linear combination of the six Ψ_i . PES measures energy differences between the molecule and its ion; any detailed interpretation of the spectra requires separate calculations for both species.

We have not yet included the effects of configuration interaction; that is, we have not carried out a spin-coupled VB calculation. However, we emphasize that in all our experience to date, this step has not altered the physical picture afforded by the spin-coupled solution alone. In the case of molecular potential energy surfaces, the addition of excited spin-coupled configurations has merely resulted in a fairly uniform lowering of the entire ground-state surface.

One of the most unacceptable facets of classical VB theory is the vast number of physically untenable ionic structures that are required. The physical reason for this failing is clear: the atomic orbitals are not allowed to distort on molecule formation. The ionic structures partially compensate for this. In the spin-coupled VB method, the orbitals are fully optimized in a flexible basis set and this problem does not arise. We have performed a small VB calculation in which all 170 possible ionic structures were included (that is, all singly, doubly and triply polar structures^{10,11}); as shown in Table 1, the total energy lowering from the inclusion of all these structures was only 0.0070 hartree. In fact, this calculation has overestimated the importance of the ionic structures: their contribution would have been even less significant if excited spin-coupled structures had been added (introducing further angular correlation).

The spin-coupled calculations have been performed using the Kotani (branching-diagram) scheme—for a discussion of different spin functions and the relationships between them see,

for example, ref. 12. We have also determined the occupation numbers of the five spin functions in the more traditional basis of Rumer functions used in classical VB theory; that is, two Kekulé structures and three Dewar structures. The results are given in Table 2. They are astonishingly close to the values given long ago by Coulson¹.

Our results suggest that the Kekulé description of benzene, as expressed in the classical VB form, is in fact much closer to reality than is a description in terms of delocalized molecular orbitals. The π -electrons of benzene are essentially localized as described by deformed atomic orbitals, and the notable aromatic characteristics of this molecule arise from the symmetric coupling of the electron spins around the carbon ring framework. This mode of description is not unfamiliar to solid-state physicists who use similar language to describe such cooperative phenomena as ferro- and antiferromagnetism. Benzene must now be considered in this light, and consequently the aromatic effects—arising from electron spin couplings—are much more profoundly quantum mechanical in origin than we have hitherto realized.

Received 8 May; accepted 25 June 1986.

1. Coulson, C. A. *Valence* 2nd edn, Ch. 9 (Clarendon, Oxford, 1961).
2. Ramsden, E. N. *A-Level Chemistry* (Stanley Thornes, 1985).
3. Gerratt, J. *Adv. atom. molec. Phys.* **7**, 141–221 (1971).
4. Gerratt, J. & Raimondi, M. *Proc. R. Soc. A* **371**, 525–552 (1980).
5. Cooper, D. L., Gerratt, J. & Raimondi, M. *Faraday Symp. chem. Soc.* **19**, 149–163 (1984).
6. Cooper, D. L., Gerratt, J. & Raimondi, M. *Adv. chem. Phys.* (in the press).
7. Dunning, T. H. & Hay, P. J. in *Methods of Electronic Structure Theory* (ed. Schaefer, H. F.) ch. 1 (Plenum, New York, 1977).
8. Pauncz, R. *The Alternant Molecular Orbital Method* (Saunders, Philadelphia, 1967).
9. Hiberty, P. C., Shaik, S. S., Lefour, J.-M. & Ohanessian, G. *J. org. Chem.* **50**, 4657–4659 (1985).
10. Norbeck, J. M. & Gallup, G. A. *J. Am. Chem. Soc.* **96**, 3386–3393 (1974).
11. Tantardini, G. F., Raimondi, M. & Simonetta, M. *J. Am. Chem. Soc.* **99**, 2913–2918 (1977).
12. Pauncz, R. *Spin Eigenfunctions* (Plenum, New York, 1979).

CO dissociation on potassium-promoted aluminium

J. Paul

Corporate Research Science Laboratories, Exxon Research and Engineering Company, Route 22 East, Annandale, New Jersey 08801, USA

Several catalytic reactions involving CO bond cleavage on transition metal surfaces are promoted by the presence of alkali atoms, and considerable interest has focused on the nature of strongly perturbed metal–CO potentials. Here I present experimental data on CO bond weakening and dissociation in carbon monoxide co-adsorbed with potassium on the (100) surface of aluminium. Atomic carbon and oxygen were observed, as well as precursor states to dissociation with CO stretch frequencies as low as $1,060\text{ cm}^{-1}$. Thus, despite its free-electron character, aluminium responds to alkali adsorption in much the same way as typical transition metals. This implies that CO bond weakening on alkali-promoted metal surfaces is caused by perturbations of the joint density of alkali and substrate sp -electrons, and that the role of d -states is not crucial. Vibrational and electronic spectra show that aluminium oxide and aluminium carbide form on annealing to 700 K. Neither CO_2 nor high-temperature CO desorption from the alkali-promoted aluminium surface was detected.

During the past few years, results from ultra-high-vacuum experiments on the co-adsorption of alkali metals and carbon monoxide on metal surfaces have been rapidly accumulating^{1–16}. Alkali additives cause an enhanced rate of CO dissociation and/or molecular scrambling, high CO desorption temperatures and a spectrum of anomalously low CO stretch frequencies. Photoemission spectra have shown considerable rearrangements

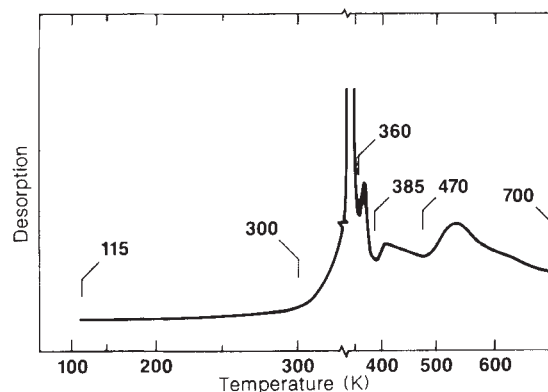


Fig. 1 Thermal desorption of potassium from Al(100) following evaporation at $\sim 115\text{ K}$. Heating rate 1.3 K s^{-1} . At the temperatures marked, the ratio (K252/Al68) of the intensity of the potassium Auger peak at 252 eV to that of the aluminium peak at 68 eV is: 115 K, 2.0; 300 K, 1.6 (the ratios at 115 K and 300 K depend on the specific potassium dose); 360 K, 0.91; 385 K, 0.84; 470 K, 0.52; 700 K, < 0.03 . The relative reduction of the K252 signal following annealing to 300 K is caused by coalescence in the potassium overlayer. (Ref. 31 and J.P., in preparation.)

of the CO π -orbitals, indicating a strongly perturbed metal–CO potential and a mutual short-range alkali–CO interaction. Similar results have been obtained for several transition metals (Cu^{1-3} , Fe^{4-6} , $\text{Ni}^{7,8}$, Pt^{9-12} , Rh^{13} and Ru^{14-16}), with different alkali adatoms ($\text{K}^{2-11,13-16}$ and $\text{Na}^{1,12}$). The cause of the anomalous potential thus appears to be largely independent of the occupation of the metal d -band and the specific electronic structure of the alkali promoter.

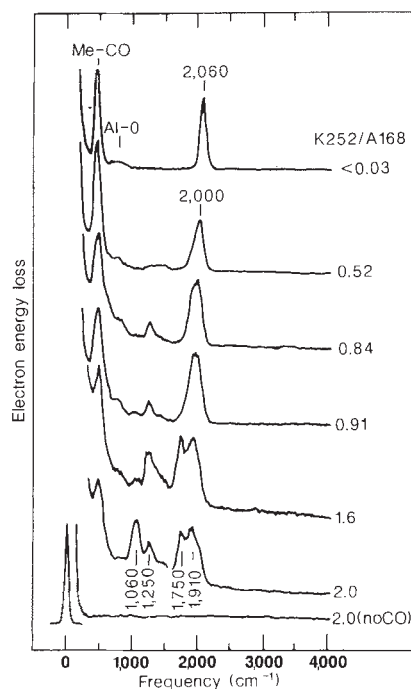


Fig. 2 Electron energy loss spectra following a fixed dose (100×10^{-6} torr s) of CO onto Al(100)/K, as a function of the potassium coverage (K252/Al68; see Fig. 1 legend). Peak positions are indicated for identified CO species. 100×10^{-6} torr s of CO from a beam-doser is equivalent to $5,000 \times 10^{-6}$ torr s from back-filling the chamber²³.

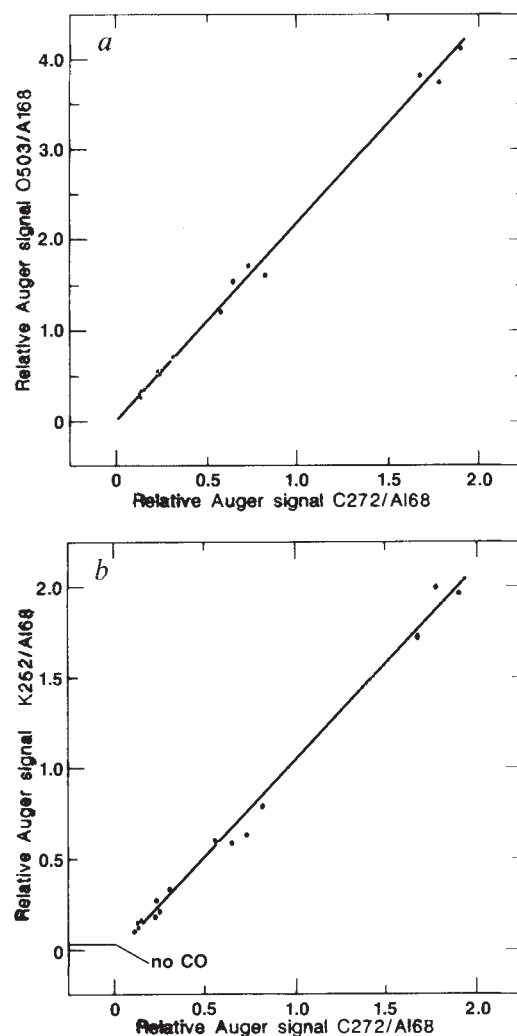


Fig. 3 Relative Auger intensities of, *a*, oxygen versus carbon and, *b*, potassium versus carbon, following co-adsorption of potassium and carbon monoxide on Al(100) and annealing to 700 K. Vibrational spectra do not indicate the presence of any molecular CO after annealing (J.P. and F.M. Hoffmann, in preparation). The lines represent least-squares fits. In *b*, 'no CO' refers to the maximum amount of potassium left after annealing of an unexposed potassium overlayer (see Fig. 1). Textbook sensitivity factors (without internal calibration) convert the slopes in *a* and *b* to the stoichiometry $K_1C_{3.5}O_{3.3}$ (ref. 33).

Calculations have addressed the response of a metal substrate to sub-monolayer coverings of partly ionized alkali atoms¹⁷⁻²¹. Despite a variety of approaches, the results are consistent with a Thomas-Fermi type of screening. Such a short-range perturbation of the density of *sp*-electrons may, however, cause a local distortion of *d*-states, which could hybridize with the orbitals of a chemisorbed CO molecule and thus cause the observed bond weakening^{5,22}.

Aluminium is of interest because its high electron density not only ensures an effective screening, with a large amplitude of the induced charge density, but also gives a work function surpassing those of alkali metals by ~ 2 eV. This implies that alkali monomers will be partly ionized at sub-monolayer coverages, much as is found with transition metal substrates. Furthermore, the simple *sp*-band character of aluminium rules out local hybridization with *d*-orbitals as a cause of anomalous CO bond weakening.

Experiments were performed in an ultra-high-vacuum chamber at 1×10^{-10} torr. The crystal mount and the 3-h *in situ* cleaning procedure have been described²³. Potassium was evaporated from a SAES getter source and CO was adsorbed from beam-dosers. Electron energy-loss (EELS) spectra were obtained at 60 cm^{-1} resolution and 5 eV primary beam energy. Finally, desorbing species were sampled with a drift-tube located 5 mm from the crystal.

Figure 1 shows potassium desorption from Al(100) following adsorption at 115 K. The spectrum resembles that of Ru(0001)/K (ref. 24), Fe(111)/K (ref. 6), and Pt(100)/K (ref. 11), which indicates that the substrate-alkali potential is insensitive to the presence of *d*-electrons. Flashing to temperatures between 115 K and 700 K followed by saturation with CO at 80 K gave the vibrational spectra shown in Fig. 2. The specific adsorption and annealing sequence of the alkali film will affect the structure of the overlayer and thus the population of CO adsorption sites. Different CO exposure and annealing sequences have identified six different CO stretch vibrations, ranging from 1,060 to 2,060 cm^{-1} (J.P. and F.M. Hoffman, in

preparation). The former value is close to the lowest reported CO stretch frequency on any metal surface²⁵. The disappearance of the potassium-modified bands (1,060–2,000 cm^{-1}) on annealing was correlated not with any desorption, but with an increased intensity near 850 cm^{-1} , which is a marker for atomic oxygen²³. I therefore assign the 1,060–2,000 cm^{-1} bands to states which are precursors to CO dissociation.

Experiments, as well as calculations, have suggested a mutual interaction between alkali atoms and carbon monoxide. The chemisorption of each species on transition metal surfaces is stabilized by the presence of the other^{11,20,26}. Two inseparable models have emerged to explain this interaction: (1) a substrate-mediated net donation of charge from alkali atoms, with charge retrieval by CO molecules^{5,9,15} and (2) a direct covalent type of interaction between alkali and CO orbitals²⁷. Whereas earlier discussions of the second model focused on bonding combinations of alkali *ns* and CO 2π orbitals, more recent work suggests that alkali (*n*–1)*p* and perturbed CO 1π orbitals may interact^{3,20,22}. Simultaneous desorption of potassium and carbon

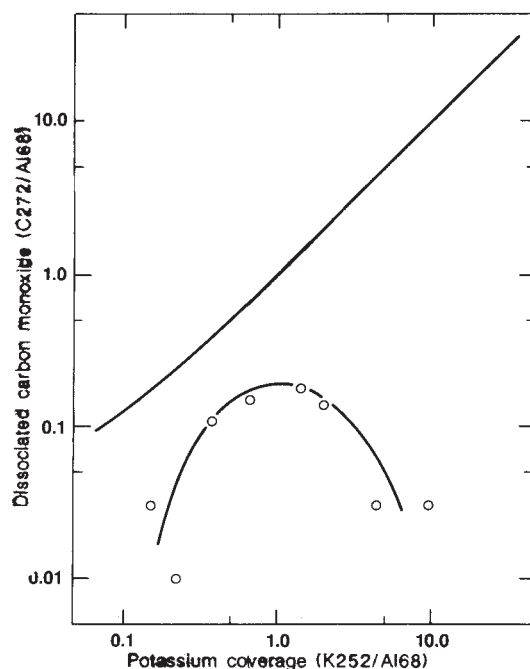


Fig. 4 Dissociated CO as a function of potassium coverage (open circles). Potassium was evaporated at 115 K without any subsequent annealing. A fixed CO dose, the equivalence of $2,500 \times 10^{-6}$ torr s, was adsorbed from a beam-doser at 80 K. The diagonal line taken from Fig. 3b represents the C:K stoichiometry of various overlayers after annealing to 700 K. Fig. 3a justifies use of the C272 signal to represent dissociated carbon monoxide.

monoxide at high temperature has been reported following co-adsorption on Ru(0001)²⁷, Fe(111)⁶ and Pt(100)¹¹. This is not observed for Al(100)/K/CO, but Auger spectra (Fig. 3) reveal a mutually strengthened chemisorption bond. Figure 3 shows a fixed stoichiometry between atomic oxygen, carbon and potassium following annealing to 700 K. The fixed oxygen to carbon ratio (Fig. 3a) demonstrates that the spectra are not affected by contaminants. The results shown in Fig. 3b point to the possible role of intermediate units of stoichiometric composition, analogous to what has been suggested for Cu(110)/K/CO (ref. 28). Vibrational spectra show that oxygen atoms occupy subsurface sites after annealing²³, which means that the driving force for aluminium oxide formation is too large to allow for any molecular desorption or recombination of CO. Auger analyses during parallel experiments with sodium promotion (J.P. and F.M. Hoffmann, in preparation) show that aluminium carbide forms on annealing to 700 K. No CO₂ desorption is observed. Electron spectroscopy measurements have revealed that potassium atoms bind to Al₄C₃ at the surface (J.P., in preparation).

The efficiency of potassium-promoted aluminium towards CO dissociation was monitored with Auger electron spectroscopy. Figure 4 shows the amount of dissociated carbon monoxide as a function of potassium coverage, following fixed low CO doses and annealing to 700 K. CO dissociates readily at intermediate potassium coverages, but hardly at all on the clean surface of bulk aluminium²⁹. Some dissociation seems to occur even at multilayer potassium coverages, but further work is needed to elucidate the importance of the morphology of coldly-deposited films^{30,31}, the CO partial pressure, and the substrate temperature during CO adsorption. The fixed C/K ratio taken from Fig. 3b may represent the maximum carbon coverage at a given potassium coverage.

Thus, potassium atoms alter the metal-CO bond on aluminium from weak to strong chemisorption, with considerable CO dissociation. The effect is similar or even stronger than on transition metal surfaces. Experiments in progress with other

alkali metals address the importance of a direct interaction between alkali $(n-1)p$ and CO 1π orbitals. Sodium and potassium atoms cause similar CO bond weakening when adsorbed on Cu(100)¹⁻³ and Pt(111)^{9,10,12}, but the binding energies of Na $2p$ electrons are >10 eV larger than those of K $3p$ electrons³². This suggests that the energy difference between alkali $(n-1)p$ and CO 1π orbitals is not crucial, as would be expected for model (2) above, and a perturbation approach to covalent bonding. Therefore, the perturbed sp -electron density is probably the main cause of the anomalously deep metal-CO potential on alkali-promoted surfaces, in combination with a minor, mutually stabilizing CO-alkali interaction mediated by the embedding electron gas.

I acknowledge discussions with F. M. Hoffmann and S. R. Keleman.

Received 25 April; accepted 31 July 1986.

1. Wallden, L. *Surf. Sci.* **134**, L513-L515 (1983).
2. Somerton, C., McConville, C. F., Woodruff, D. P., Grider, D. E. & Richardson, N. V. *Surf. Sci.* **138**, 31-39 (1984).
3. Heskett, D., Strathy, I., Plummer, E. W. & dePaola, R. A. *Phys. Rev.* **B32**, 6222-6237 (1985).
4. Broden, G., Gafner, G. & Bonzel, H. P. *Surf. Sci.* **84**, 295-314 (1979).
5. Benziger, J. & Madix, R. J. *Surf. Sci.* **94**, 119-153 (1980).
6. Seip, U., Bassignani, I. C., Küppers, J. & Ertl, G. *Surf. Sci.* **160**, 400-418 (1985).
7. Lee, J., Arias, J., Hanrahan, C. P., Martin, R. M. & Metiu, H. *Phys. Rev. Lett.* **51**, 1991-1994 (1983).
8. Whitman, L. J. & Ho, W. *J. chem. Phys.* **83**, 4808-4816 (1985).
9. Crowell, J. E., Garfunkel, E. L. & Somorjai, G. A. *Surf. Sci.* **121**, 303-320 (1982).
10. Kiskinova, M., Pirug, G. & Bonzel, H. P. *Surf. Sci.* **133**, 623-629 (1983).
11. Bertolini, J. C., Delichere, P. & Massardier, J. *Surf. Sci.* **160**, 531-541 (1985).
12. Sette, F. *et al. Phys. Rev. Lett.* **54**, 935-938 (1985).
13. Crowell, J., Tysoe, W. T. & Somorjai, G. A. *J. phys. Chem.* **89**, 1598-1601 (1985).
14. Hoffman, F. M. & dePaola, R. A. *Phys. Rev. Lett.* **52**, 1697-1700 (1984).
15. Weimer, J. J., Umbach, E. & Menzel, D. *Surf. Sci.* **159**, 83-107 (1985).
16. Madey, T. E. & Benndorf, C. *Surf. Sci.* **164**, 602-624 (1985).
17. Joyner, R. W., Pendry, J. B., Saldin, D. K. & Tennyson, S. R. *Surf. Sci.* **138**, 84-94 (1984).
18. Lang, N. D., Holloway, S. & Nørskov, J. K. *Surf. Sci.* **150**, 24-38 (1985).
19. Feibelman, P. J. & Hamann, D. R. *Surf. Sci.* **149**, 48-66 (1985).
20. Wimmer, E., Fu, C. L. & Freeman, A. J. *Phys. Rev. Lett.* **55**, 2618-2621 (1985).
21. Paul, J. & Rosén, A. *Surf. Sci.* **127**, L93-L97 (1983).
22. Eberhardt, W. *et al. Phys. Rev. Lett.* **54**, 1856-1859 (1985).
23. Paul, J. & Hoffmann, F. M. *J. phys. Chem.* (in the press).
24. dePaola, R. A., Hrbek, J. & Hoffman, F. M. *J. chem. Phys.* **82**, 2484-2498 (1985).
25. Zaera, F., Kollin, E. & Gland, J. L. *Chem. Phys. Lett.* **121**, 464-468 (1985).
26. Hoffman, F. M., Hrbek, J. & dePaola, R. A. *Chem. Phys. Lett.* **106**, 83-86 (1984).
27. Paul, J. *Prog. Colloid Polym. Sci.* **70**, 57-61 (1985).
28. Lackey, S., Surman, M., Jacobs, S., Grider, D. & King, D. A. *Surf. Sci.* **152/153**, 513-521 (1985).
29. Paul, J. & Hoffman, F. M. *Chem. Phys. Lett.* (in the press).
30. Albano, E. V. *et al. Phys. Rev. Lett.* **51**, 2314-2317 (1983).
31. Lindgren, S. Å., Paul, J., Wallden, L. & Westrin, P. J. *Phys. C* **15**, 6285-6291 (1982).
32. Surnev, L., Bliznakov, G. & Kiskinova, M. *Solid St. Commun.* **37**, 87-90 (1981).
33. Davis, L. E., MacDonald, N. C., Palmberg, P. W., Riach, G. E. & Weber, R. E. (eds) *Handbook of Auger Electron Spectroscopy* (Perkin-Elmer Corporation, Eden Prairie, 1978).

Icosahedral symmetry carbon cage molecules

D. J. Klein, W. A. Seitz & T. G. Schmalz

Theoretical Chemical Physics Group, Department of Marine Sciences, Texas A&M University at Galveston, Galveston, Texas 77553, USA

Graphite, when vaporized by laser irradiation, produces a very stable cluster of 60 carbon atoms which, it has been suggested, takes the uniquely elegant form of one of the archimedean semi-regular polyhedra—the truncated icosahedron^{1,2}. In considering the possibility of other structures of especial stability, we are led by a sequence of fairly general arguments to consider a relatively restricted family of high-symmetry cage structures, which correspond to novel convex polyhedra of icosahedral symmetry. Considering each of the smaller (computationally accessible) species in the family we predict here that C₂₀ is unstable; C₈₀ and C₁₄₀ have moderate resonance energies but are open shell; and C₆₀, C₁₈₀ and C₂₄₀ are closed shell especially stable forms. The latter two are possible stable carbon cages not yet experimentally characterized.

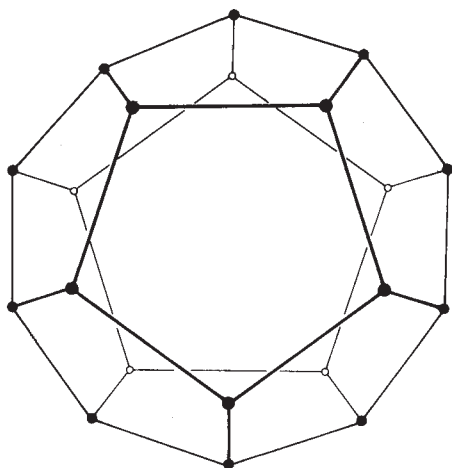


Fig. 1 The dodecahedron on whose faces 12 copies of any one of the cells of Fig. 2 may be placed.

Stable structures should be those for which every carbon atom attains its tetravalency, preferably with little strain. This may be accomplished with a linear poly-yne chain bent to close into a ring^{3,4}, but limitation of the carbon-carbon σ -bond strain implicates larger rings which have little π -resonance energy. Significant resonance stabilization may occur, however, in planar frameworks bent around (in two directions) to form cages with three σ -bonds and one π -bond incident on each site. To maximize local stability there should be little local curvature, so that the σ -skeleton achieves most nearly the planar sp^2 ideal and also that the 'overlap' between the remaining neighbour π -orbitals is as large as possible. Thus, the cages we consider all correspond to convex polyhedra.

For such polyhedra, the constraints of Euclidean space require that

$$3f_3 + 2f_4 + f_5 = 12 + f_7 + 2f_8 + 3f_9 + 4f_{10} + \dots \quad (1)$$

where f_n denotes the number of n -sided faces. Thus cage formation necessitates incorporation of some five, four or three membered rings. In seeking the most stable conjugated structures we note three general chemical observations: (1) hexagonal (benzenoid) rings have the greatest (aromatic) stability; (2) quadrilateral (and to some extent octagonal) rings have very reactive π -electron systems; (3) triangular rings are unstable due to very high angular strain. As a consequence we take $f_3 = f_4 = 0$ and eliminate $n \geq 7$ rings in equation (1) so we are limited to polyhedra having only hexagonal and 12 pentagonal faces.

These structures already correlate with three general experimental observations: (1) only species with an even number of carbon atoms are observed^{1,6,7} (at long times); (2) the large cluster ionization potentials (< 5.0 eV) are similar to various graphites (3.9–4.9 eV) (short poly-yne's have ionization potentials near 10 eV)^{4,6}; (3) in low-temperature experiments all species with an even number of C atoms > 30 are observed—consistent with the fact that polyhedra meeting our criteria exist for all even numbers of vertices, $v > 24$ (ref. 8).

Turning to the question of long-term stability, we limit candidates to those with high symmetry. Since effective (screened) interactions are dominantly local, repetition of an especially stable local structure should lead to an overall global structure of especial stability. But such repetition, if feasible under the constraints of realization in euclidean space, often leads to a system of high symmetry. Icosahedral symmetry (as in the proposed C_{60} structure) is the highest such (point group) symmetry and leads to near-spherical cages, thereby minimizing strain locally.

It can be proved that the only icosahedral cage structures that can be constructed form an infinite sequence with the number of vertices given by⁹

$$v = 20(h^2 + hk + k^2), \quad 0 \leq k \leq h \quad (2)$$

with h and k integers. These cage structures are conveniently viewed as being built up from several symmetry-equivalent fragments or cells. One choice for these building-block cells corresponds to the 12 faces of a dodecahedron, as depicted in Fig. 1. That is, a pattern of sites and edges is made on each cell so that when 12 copies of such a cell are properly adjoined, a desired cage structure results. The six simplest cells are depicted in Fig. 2. The first (for C_{20}) corresponds to the dodecahedron, one of the five regular (platonic) solids. The second (for C_{60}) is a truncated icosahedron, the semi-regular (archimedean) form dubbed Buckminsterfullerene by Kroto *et al.*¹. The C_{80} and C_{140} structures can be obtained by truncation of the five-valent vertices of previously described¹⁰ duals of two semi-regular polyhedra: the icosidodecahedron and the snub-cuboctahedron. The C_{180} and C_{240} structures can be obtained from truncation of the duals to the truncated icosahedron (C_{60}) and the C_{80} structure, respectively. The dual to the C_{140} structure is identified with some viruses^{9,11}. The duals to other higher structures (usually with $k = 0$) are associated with many of Buckminster Fuller's geodesic domes¹². The icosahedral crystallites of Mackey¹³, apparently realized as gold clusters¹⁴, also involve such duals (with $k = 0$). Teo and Sloane have described 'dodecahedral' crystallites corresponding to the $h = k$ duals¹⁵.

Quantitative estimates of the stability of the selected molecules are obtained through ordinary Hückel molecular orbital theory

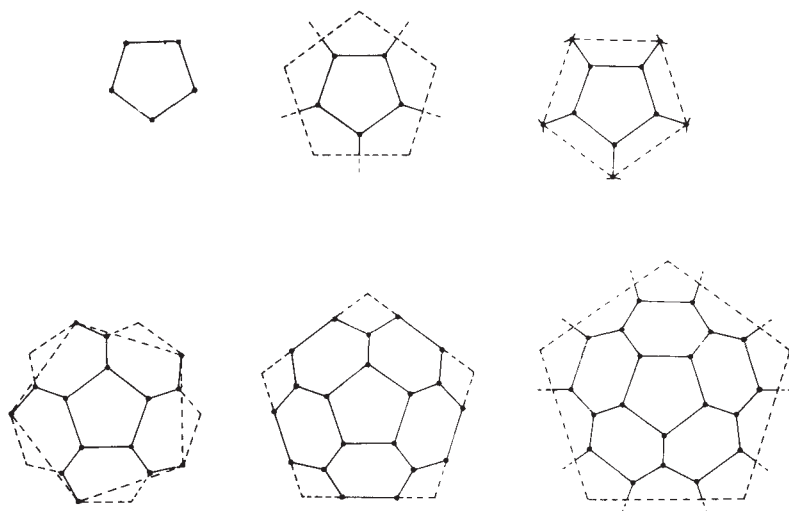


Fig. 2 The dodecahedral-face cells of the six smallest polyhedra of our preferred class.

Table 1 Structural characteristics

	C ₂₀	C ₆₀	C ₈₀	C ₁₄₀	C ₁₈₀	C ₂₄₀
K	36	12,500	140,625	2,167,239,697	1,389,029,765,625	21,587,074,966,666,816
# ⁽⁶⁾	0	83,160	835,800	34,362,948,480	33,310,430,346,840	701,366,195,951,634,560
# ⁽⁸⁾	240	0	0	0	0	0
# ⁽¹⁰⁾	120	59,760	644,280	31,487,794,920	40,033,439,580,420	642,190,547,816,179,080
cos $\langle \theta^2 \rangle^{1/2}$	0.766	0.920	0.940	0.965	0.973	0.980

Table 2 Resonance energies

	Simplest MO and VB methods		$E(\text{HOMO}) - E(\text{LUMO})$ β	Refined MO and VB methods	
	Kekulé count $-\Delta\epsilon/J$	Simple HMO $\Delta\epsilon/\beta$		Hess and Schaad $\Delta\epsilon/\beta$	Conjugated circuit $-\Delta\epsilon$ (eV) (Herndon's parameters)
Benzene	0.131	0.333	2.000	0.065	0.140
Pentacene	0.092	0.388	0.439	0.038	0.084
Coronene	0.141	0.440	1.078	0.053	0.146
C ₆₀	0.144	0.485	0.507	-0.037	0.005
C ₁₂₀	0.203	0.488	0.351	-0.034	-0.072
C ₂₀	0.203	0.471	0	-0.051	-0.031
C ₆₀	0.178	0.553	0.757	0.031	0.120
C ₈₀	0.168	0.544	0	0.022	0.080
C ₁₄₀	0.174	0.560	0	0.038	0.130
C ₁₈₀	0.176	0.567	0.580	0.045	0.166
C ₂₄₀	0.178	0.569	0.496	0.047	0.155
Graphite	0.183	0.575	0	0.053	0.168

(HMO), a refined method developed by Hess and Schaad, and by quantitative resonance theory. In ordinary HMO theory, total π -electron energy is referenced against a set of non-resonant ethylenes leading to the (unreasonable) prediction of high stability for all large carbon cages. The expression for the resonance energy (RE) through the more accurate Hess and Schaad method reduces to¹⁶

$$(RE) = E_{\pi}(\text{HMO}) - 1.5216 \nu\beta \quad (3)$$

where β is the usual Hückel parameter. Results from these types of molecular orbital (MO) methods, however, depend (as in equation (3)) on differences between two large numbers, so that small differences in adjustable parameters lead to large relative differences in resonance energy.

Direct calculation of resonance energy is obtained through Herndon-Simpson^{17,18} or conjugated circuit^{19,20} resonance theory. In this approach, the resonance energy is given as the ratio H/K where K is the total number of Kekulé structures, and

$$H = \sum_{m \geq 1} (R_m \#^{(4m+2)} + Q_m \#^{(4m)}) \quad (4)$$

where the R_m and Q_m are oppositely-signed parameters which decrease in magnitude (near geometrically) with increasing m and where $\#^{(2m)}$ is the sum of the number of conjugated $2m$ -circuits in the various Kekulé structures. (A conjugated $2m$ -circuit is a length- $2m$ cycle with alternating single and double bonds). We enumerate the Kekulé structures through a graph-theoretic transfer matrix technique²¹, which with slight modification may also be used to count the conjugated circuits. Alternative parameterizations of equation (4) are available, but in contrast to Hückel theory, this expression for the resonance energy does not involve a difference between large numbers, and different parameter sets give rather similar answers.

Exact graph-theoretic (structural) results, K and nonzero $\#^{(2m)}$ with $m \leq 5$, are reported in Table 1 for the six icosahedral cages. For later comparison like calculations were made for the C₁₂₀ cage structure recently studied by Haymet using Hückel

theory²². This cage (an archimedean semi-regular polyhedron, the great rhombicosidodecahedron), has numerous 4- and 10-membered rings. As a further example, we treated the remaining semi-regular solid with both icosahedral symmetry and three edges incident at each vertex; this is the truncated dodecahedron giving a C₆₀ cage with 3- and 10-membered rings.

The computational schemes utilized here make several approximations. The Hückel approach neglects explicit correlation effects, while the resonance-theory approach neglects explicit inclusion of higher configurations. All of our schemes neglect variations in bond length (due to π -bond localization) and coupling between σ - and π -systems. Modifications of the model parameters away from their values for planar skeletons, however, may be estimated approximately. Let θ denote the angle of deviation of two near-neighbour π -orbitals away from being parallel. Then the dominantly one-electron resonance-integral parameter of Hückel theory should scale as $\cos \theta$ whereas the dominantly two-electron exchange-integral parameter of resonance theory should scale as $\cos^2 \theta$. Table 1 also lists $\cos \langle \theta^2 \rangle^{1/2}$ for each cage considered, which may be applied as a correction factor to the resonance energies.

The uncorrected π -electron resonance energies per site are reported in Table 2. For C₆₀ several previous calculations have been given²³⁻²⁹, and several others are in progress. In addition, we also report a quantification of the standard idea that the count of Kekulé structures gives an estimate of the resonance energy²⁵. C₂₀ and the two semi-regular icosahedral-symmetry structures falling outside of our 'preferred' class are predicted to be quite unstable, at least with the procedures believed to be more reliable (especially the conjugated-circuit approach). There is little support for the suggestion²² that the 30 four-membered rings in C₁₂₀ might be stabilized in the overall structure.

However, the five remaining icosahedral-symmetry cages (in our 'preferred' class) are all indicated to have some stabilizing π -resonance energy. The C₈₀ cage is the least stable of these, and, along with C₁₄₀, has a HOMO-LUMO (highest occupied

MO—lowest occupied MO) gap of 0. This feature may be taken as an indicator of reactivity for these two structures. The three remaining species C_{60} and especially C_{180} and C_{240} are predicted to be quite stable. While the method of Hess and Schaad is a significant improvement over standard MO theory, we think that the conjugated circuit approach provides the best estimate of resonance stabilization for these carbon cage systems.

This research was supported by the Robert A. Welch Foundation of Houston, Texas.

Received 19 May; accepted 1 August 1986.

1. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **121**, 33–37 (1985).
2. Liu, Y. *et al. Chem. phys. Lett.* **126**, 215–217 (1986).
3. Pitzer, K. S. & Clementi, E. *J. Am. chem. Soc.* **81**, 4477 (1959).
4. Ewing, D. W. & Pfeiffer, G. V. *Chem. phys. Lett.* **86**, 365 (1982).
5. Grunbaum, B. *Convex Polytopes* Ch. 13 (Interscience, New York, 1967).
6. Rohlfing, E. A., Cox, D. M. & Kaldor, A. *J. chem. Phys.* **81**, 3322–3330 (1984).
7. Bloomfield, L. A., Geusic, E. M., Freeman, R. R. & Brown, W. L. *Chem. phys. Lett.* **121**, 33–37 (1985).
8. Grunbaum, B. & Motzkin, T. S. *Can. J. Math.* **15**, 744–751 (1963).
9. Caspar, D. L. D. & Klug, A. *Cold Spring Harbor Symp. on Quantitative Biology* **27**, 1–24 (1962).
10. Johnson, N. W. *Can. J. Math.* **18**, 169–200 (1966).
11. Klug, A. & Finch, J. T. *J. molec. Biol.* **11**, 403–423 (1965).
12. Marlin, W. (ed.) *The Artifacts of R. Buckminster Fuller—A Comprehensive Collection of His Designs and Drawings* (Garland, New York, 1984).
13. Mackey, A. L. *Acta crystallogr.* **15**, 916 (1962).
14. Iijima, S. & Ichihashi, T. *Phys. Rev. Lett.* **56**, 616–619 (1986).
15. Teo, B. K. & Sloane, N. J. A. *Inorg. Chem.* **24**, 4545–4558 (1985).
16. Hess, B. A. & Schaad, L. J. *J. Am. chem. Soc.* **93**, 305–310 (1971).
17. Herndon, W. C. *J. Am. chem. Soc.* **95**, 2404–2406 (1973).
18. Herndon, W. C. & Ellzey, M. L. Jr *J. Am. chem. Soc.* **96**, 6631–6642 (1974).
19. Randic, M. *Tetrahedron* **31**, 1477–1481 (1975).
20. Randic, M. & Trinajstić, N. *J. Am. chem. Soc.* **106**, 4428–4434 (1984).
21. Klein, D. J., Schmalz, T. G. & Hite, G. E. *J. computational Chem.* **7**, 443–456 (1986).
22. Haymet, A. D. J. *Chem. phys. Lett.* **122**, 421–424 (1985).
23. Bochvar, D. A. & Gal'pern, G. E. *Doklady Akad. Nauk SSSR* **209**, 610–612 (1973).
24. Haymet, A. D. J. *J. Am. chem. Soc.* **108**, 319–321 (1986).
25. Klein, D. J., Schmalz, T. G., Seitz, W. A. & Hite, G. E. *J. Am. chem. Soc.* **108**, 1301 (1986).
26. Haddon, R. C., Brus, L. E. & Raghavachari, K. *Chem. phys. Lett.* **125**, 459–464 (1986).
27. Disch, R. L. & Schulman, J. M. *Chem. phys. Lett.* **125**, 465–466 (1986).
28. Newton, M. D. & Stanton, R. E. *J. Am. chem. Soc.* **108**, 2469–2470 (1986).
29. Fowler, P. W. & Woolrich, J. *Chem. phys. Lett.* **127**, 78–83 (1986).

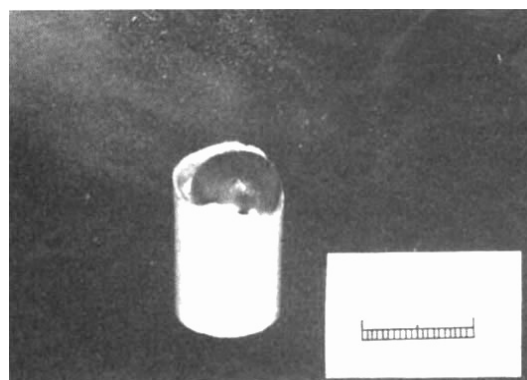


Fig. 1 Burning pressed mixture of tetrazole and sodium tetrazolate. The ILSF is in the upper part of the cylinder; over the cylinder, smoke and dispersed products of burning. Scale bar, 2 cm.

Sodium tetrazolate (structure (2)) was obtained by interaction of equimolar amounts of sodium ethylate and tetrazole and subsequent removal of the ethanol *in vacuo*. The substances were identified by X-ray and chromatographic analysis.



The sample mixtures were compacted at a pressure of $6.8 \times 10^8 \text{ N m}^{-2}$ to form 8-, 10-, 15- and 20-mm cylinders, 30 mm in height. Combustion was initiated by a red-hot Ni/Cr alloy filament, in a nitrogen atmosphere and in air.

Pure tetrazole ignites and burns readily in a nitrogen atmosphere, producing a slightly luminous flame due to the decomposition of vapours evaporated from the surface of the molten layer, whose temperature reaches 670 K (measured with a manganin/constantan thermocouple with a 10- μm thick junction). The rate of combustion at atmospheric pressure was $\sim 0.45 \text{ mm s}^{-1}$. On contact with the air, the products of oxidation of the vapours and decomposition of tetrazole may be superimposed on the decomposition flame. The samples of pure tetrazole burn completely to leave no solid residue.

Adding sodium tetrazolate to tetrazole changes the combustion pattern—explosive-like sparking can be seen in the molten pre-flame layer. With 35–38 wt % sodium tetrazolate in the mixture, the combustion pattern becomes extraordinary. When samples of this mixture begin to burn, a molten layer is produced, which gives off gaseous decomposition products. This is followed by an incandescent liquid spheroidal formation (ILSF), 1–2 mm in diameter and light red in colour, which moves quickly on the melt surface. According to qualitative investigations, carbon and iron intensify the ILSF formation. Following ILSF formation, the combustion becomes a self-sustained process. The ILSF quickly moves over the surface covered by the melt, grows in diameter to 13–17 mm for every 20 mm of the sample (Fig. 1) and then stops growing and moving quickly. The combustion surface acquires a concave shape. The ILSF governs the combustion process; there is no gas flame. White vapours, which condense on cooling, form on the ILSF edges that are in contact with the initial substance, and the reaction products, which are the same as those in a cooled ILSF, are dispersed (X-ray analysis data). The ILSF temperature, measured by introducing a W/Re microthermocouple with 50- μm wire directly into the ILSF, was 1,030 K.

The ILSF can be made to move on the sample surface in the desired direction by tilting the sample or by changing the ILSF using a gas flow. As a result, the combustion process can, to a certain extent, be controlled. For instance, combustion can be terminated, by turning the sample over or by blowing the ILSF off the combustion surface with an air flow. (Occasionally, a

A new type of self-organization in combustion

A. I. Lesnikovich, V. V. Sviridov, G. V. Printsev,
O. A. Ivashkevich & P. N. Gaponik

Institute of Physico-Chemical Problems,
Byelorussian State University, Minsk 220080, USSR

We report here that a pressed mixture of tetrazole and sodium tetrazolate can burn, both in an inert atmosphere and in the air, forming an incandescent liquid spheroidal reaction zone of definite size. This and other features of the reaction zone lead to its description as a liquid flame, in which peculiar chemical processes can occur.

Combustion is a self-organized process¹. Processes of unsteady burning—self-oscillation and spin combustion—have been recognized². Most condensed explosive substances burn at the expense of their own components and do not require an oxidizer. Such burning usually comprises the evaporation of the basic substance, or its partial decomposition, in the condensed phase, followed by burning of the gas formed. Burning may also take place in the condensed phase only—this is a solid-phase gasless combustion³ which proceeds with the formation of a melt or a flameless gas^{3,4}. Flameless combustion has recently been described for several tetrazole compounds⁵.

When mixed with sodium tetrazolate, tetrazole burns in the condensed phase. This type of combustion has several interesting features.

Tetrazole (structure (1)) was obtained by the technique described in ref. 6 and was purified by sublimation *in vacuo*.

new ILSF may form spontaneously on the combustion surface, to evolve like the previous one.)

An ILSF of stable size formed on 20-mm cylinders. With samples of smaller diameter, the ILSF rolled down the combustion surface before its stable size had been reached.

Hardening the ILSF on a cold surface yields a light-grey, porous ingot. The ingot mass is independent of the mass of the burnt portion of the sample: it is determined by the size of the ILSF when it hardens. The density of the ILSF, calculated from the ingot mass and the ILSF size measured from photographs, was $\sim 0.1 \text{ g cm}^{-3}$, which is higher by two orders of magnitude than that of gases and lower by an order of magnitude than that of liquids. Obviously, the liquid in the ILSF contains gas bubbles.

The infrared spectrum of the ingot substance (Specord-75IR, KBr pellet) is comparatively simple, comprising broadened high-intensity bands at $1,381 \text{ cm}^{-1}$ (probably $\nu_{\text{N}=\text{N}-}$), $1,542 \text{ cm}^{-1}$ ($\nu_{\text{C}=\text{N}-}$ in conjugated polycyclic systems), $2,151 \text{ cm}^{-1}$ ($\nu_{\text{N}=\text{C}=\text{N}-}$) and a low-intensity band at $2,218 \text{ cm}^{-1}$ ($\nu_{\text{C}\equiv\text{N}}$). Bands at 842 and $1,417 \text{ cm}^{-1}$ reveal the presence of sodium carbonate, which has formed on the ILSF surface in air.

According to the elemental analysis data (Hewlett-Packard CHN-analyser), the atomic constitution of the organic part of the ingot substance corresponds to the relation: C:H:N = 4.77:1.00:5.63. Crystalline sodium carbonate, constituting 50–52% of the starting ingot weight, remains in the quartz cell after burning the ingot in the air at $1,120 \text{ K}$.

Thermal analysis of the ingot substance, carried out in an argon atmosphere (Paulik, Paulik, Erdey derivatograph, MOM Hungary; heating rate, 10 K min^{-1} ; sample weight, 55 mg), reveals an endothermal event, corresponding to a 9% weight loss, on the differential thermal analysis (DTA) curve at 700 K . A broadened exothermal event can be seen on the DTA curve at 890 K , at which temperature the sample begins to lose weight. By the end of the experiment ($1,170 \text{ K}$) the sample has undergone a weight loss of 33%.

The data indicate that sodium salts and polycyclic nitrogen-carbon compounds with high thermal stability are synthesized within the ILSF volume during combustion of the mixture in air and in a nitrogen atmosphere.

Mass spectrometric analysis shows that nitrogen is the main gaseous combustion product. The peak with mass to charge ratio $m/e = 44$ is the second most intense (probably diaziridine, CN_2H_4). An HCN peak may also be present in the spectrum.

The above type of combustion is angular in many respects. The chemistry of the combustion process is unusual: the tetrazole and sodium tetrazolate do not decompose to the starting materials used in their synthesis ($\text{HN}_3 + \text{HCN}$); heat-resistant sodium salts and polycyclic nitrogen-carbon compounds are synthesized concurrently with the decomposition of the initial substances. Furthermore, the large volume of tetrazole in the mixture does not produce the characteristic gas-phase decomposition flame. The combustion process is localized in the ILSF and on its boundary with the initial substance. The ILSF probably dissolves the initial components and absorbs the tetrazole vapours. The formation of the ILSF itself is remarkable, as is the cessation of its growth that testifies to the equality of reagent input and product output. The reaction products may be released by evaporation, and the dispersion from the surface of contact between the ILSF and the initial substance. The high-temperature conditions in the ILSF are also unusual, as tetrazole and sodium tetrazolate are unstable at temperatures above 590 K . These conditions are made possible by the fast exothermal decomposition of the initial substance which continuously feeds the ILSF. This type of combustion differs from the self-propagating decomposition described earlier^{3–5} in the high temperatures attained and the peculiar features of the zone responsible for the combustion. The temperature of the initial decomposition is almost twice as high, and allows new chemical reactions to proceed.

With a further increase in the sodium tetrazolate content of the mixture, the combustion process is also accompanied by the formation of an incandescent liquid zone of intensive transformation. The size of this zone, however, grown indefinitely, and the substance rolls down the burning surface; thus, no zone having a stable size is formed. Similar observations are obtained during the combustion of pure sodium tetrazolate.

The basic combustion processes in the above examples occur in the zones of intensive transformation, the specific features of which determine the entire combustion pattern. If we ignore the fact that these zones represent a liquid in which a complex set of reactions take place, the entire process resembles the combustion of substances, including tetrazole, which produce a decomposition flame. Such flames transform most of the substance and release most of the heat necessary for flame luminescence and for the required heat flow to the basic substance. All this occurs in the zones of intensive transformation; therefore, they may be referred to as liquid flame zones. This term does not seem to be innovative; in fact, flame propagation has been mentioned in the case of gasless combustion in the solid phase. In general, this type of combustion is called a deflagration wave.

Thus the combustion of sodium tetrazolate and, in particular, the ILSF may be regarded as a self-organized liquid-flame chemical reactor. Such a reactor features a peculiar inter-relationship between three aggregate states of the substance that cannot be observed during solid-phase combustion, combustion accompanied by the formation of a melt of combustion products, or combustion accompanied by the formation of a molten layer of the initial substance. In particular, the influence of the inert gas pressure on the rate of combustion of the sodium tetrazolate and tetrazole mixture (when the pressure was increased from 2 to 6 Mpa the burning rate increased from 2.4 to 8.3 mm s^{-1}) may be attributed to its effect on the shape and size of the liquid flame, on the flame's velocity and type of motion on the initial substance surface, on the solubility of the initial products in it, and the removal of the end products. The liquid flame also allows substances to remain in the combustion zone for a long time. As a result, comparatively slow endothermal and exothermal processes, in particular polymerization, may take place in a liquid-flame reactor.

Received 19 February; accepted 2 July 1986.

1. Zeldovich, Ja. B., Librovich, V. B. & Merzhanov, A. G. *Priroda* **2**, 3–43 (1985).
2. Merzhanov, A. G. *Arch. Combust.* **1**, 23–48 (1981).
3. Gorbunov, V. V. & Shidlovskij, A. A. *Phys. Goren. Vzryva* **6**, 471–474 (1970).
4. Lesnikovich, A. I., Levchik, S. V., Kovalenko, K. K. & Guslev, V. G. *Thermochim. Acta* **81**, 245–260 (1984).
5. Lesnikovich, A. I. et al. *Doklady Acad. Nauk belorussk.* **29**, 825–827 (1985).
6. Japan Patent 7247031, *Chemical Abstr.* **78**, 111331 (1973).

Isotopic composition of osmium in terrestrial samples determined by accelerator mass spectrometry

U. Fehn*†, R. Teng*†, D. Elmore† & P. W. Kubik†

* Department of Geological Sciences, University of Rochester, Rochester, New York, 14627, USA

† Nuclear Structure Research Laboratory, University of Rochester, Rochester, New York, 14627, USA

Interest has revived in the rhenium/osmium system due to significant advances in detection methods^{1–4} and because of its use for the identification of extraterrestrial material at the Cretaceous/Tertiary boundary⁵. Another, perhaps more important application is its use as a tracer for the origin of crustal material. This application is, however, severely restricted by the very low concentrations typical of crustal rocks. We present here the first results for the isotopes ¹⁸⁷Os, ¹⁸⁸Os, and ¹⁸⁹Os measured in sub-p.p.b. (parts per 10⁹) level samples by accelerator mass spec-

Table 1 Comparison of osmium ratios in the Canyon Diablo iron meteorite

Material	$^{188}\text{Os}/^{189}\text{Os}^*$	$^{187}\text{Os}/^{189}\text{Os}$	$^{187}\text{Os}/^{186}\text{Os}$
Canyon Diablo†	0.8175 ± 0.0003	0.109 ± 0.0003	1.114 ± 0.003
Canyon Diablo‡	0.791 ± 0.018	0.101 ± 0.004	1.025 ± 0.040 §

For each measurement, ~5 mg of crushed meteorite were placed in the Cs sputter source. The two analyser settings and the terminal voltage were manually adjusted to select masses 187, 188, and 189 which were counted for 1–10 min each. Mass 186 was not measured due to interference from ^{186}W . Because the $^{187}\text{Os}/^{186}\text{Os}$ ratio is customarily used in the literature we also give this ratio which is calculated from the measured $^{187}\text{Os}/^{189}\text{Os}$ ratio and the ratio for $^{189}\text{Os}/^{186}\text{Os}$ of 10.15 (ref. 10). The data analysis was carried out as reported previously¹⁵, but ratios were not normalized to the Nier values. Standard deviations (1σ) are given here and in Table 2.

* Nier's value for $^{188}\text{Os}/^{189}\text{Os} = 0.826$ (ref. 14).

† From ref. 10, normalized to Nier's values for $^{192}\text{Os}/^{188}\text{Os}$.

‡ This study.

§ Calculated.

metry (AMS). Our first exploratory measurements on extraterrestrial and crustal material with Os concentrations between 2 p.p.m. (parts per 10^6) and 0.006 p.p.b. show a detection limit of ~0.01 p.p.b. and a precision of better than 10%. The $^{187}\text{Os}/^{186}\text{Os}$ ratios measured for samples for the Canyon Diablo meteorite and from the East Clearwater crater in Quebec, Canada, were close to 1, typical of extraterrestrial material. In contrast, ratios in samples from the Ries crater in Germany were around 10, which is evidence of predominantly crustal osmium in these rocks.

Crustal concentrations of osmium are ~0.01 p.p.b., and are about two and four orders of magnitude below those in mantle-derived rocks and in extraterrestrial material, respectively^{1,6–8}. Although rhenium and osmium are both depleted in the crust, the degree of depletion is significantly higher for osmium than for rhenium, by factors of 100 or more^{1,9}. The preferential depletion of osmium, combined with the decay of ^{187}Re into ^{187}Os (half-life $\sim 4.5 \times 10^{10}$ yr)^{10,11} causes a significant enrichment of ^{187}Os relative to the other osmium isotopes in crustal rocks. While the $^{187}\text{Os}/^{186}\text{Os}$ ratios in the mantle and in extraterrestrial material are very close to unity, ratios in crustal material are expected to fall between 10 and 30 (refs 1, 8). This difference in isotopic signature is the basis for the applications of this system as tracer for the origin of crustal material and the distinction of extraterrestrial material^{1,5,8,12,13}. Because of these characteristics (very low concentrations of osmium, but very large differences in isotopic signatures) a method such as AMS with its high sensitivity and moderate accuracy is well suited for the osmium system.

To test the AMS method we used unprocessed samples from the Canyon Diablo iron meteorite. The osmium and rhenium concentrations in this meteorite are 2 and 0.2 p.p.m., respectively, and the $^{187}\text{Os}/^{186}\text{Os}$ ratio is 1.11 (ref. 10). The results for this meteorite (Table 1) are close to, but slightly lower than the ratios determined by previous workers^{10,14}. If the values are corrected for mass fractionation the $^{187}\text{Os}/^{186}\text{Os}$ ratio changes to 1.117, in very close agreement with the normalized values of Luck and Allègre¹⁰. The ratios obtained in seven individual measurements of five splits of this meteorite show good reproducibility (Fig. 1).

An important result of these measurements is the absence of significant isobaric interference from ^{187}Re . In contrast to counting rates between 1,000 and 2,000 c.p.m. (counts per minute), typical for mass 187, rates for ^{185}Re were found to be between 3 and 5 c.p.m. Counting rates higher by factors between 30 and 50 were, however, found at mass 201, which indicates that rhenium forms negative ions preferentially as an oxide. A comparison of the counting rates for rhenium and osmium suggests that negative monoatomic rhenium ions are formed at a rate lower by a factor of ~1,000 than the corresponding osmium

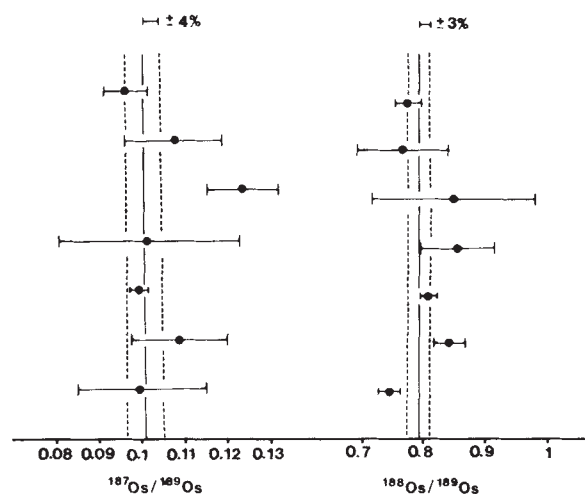


Fig. 1 Individual results for the seven runs of the Canyon Diablo meteorite. Each point represents data from at least three cycles. Lines indicate the weighted average and the standard deviation (1σ). The methods for measuring the osmium isotopes are similar to those reported for ^{129}I (ref. 16), but the settings of injector magnet, terminal voltage, and electrostatic analyser were changed to select the desired osmium isotope. The isotopes were measured at charge state 5 and 20 MeV in a time-of-flight detector to ensure separation of neighbouring masses. Because the abundance ratios of the osmium isotopes are near unity (compared to ratios of 10^{-12} for $^{129}\text{I}/^{127}\text{I}$) there was never more than one isotope detected at any one setting of the analysers. Since molecules are not bound in charge state 5, the main advantage of AMS over conventional mass spectrometry is the total elimination of molecular interference.

ions. The preferential ionization of osmium in the ion source eliminates the need for a chemical separation of rhenium from osmium in most terrestrial samples.

Counting rates observed for the osmium isotopes in Canyon Diablo were about two orders of magnitude higher than necessary to measure ratios with accuracies below 10%. It can thus be expected that samples with osmium concentrations above 20 p.p.b. could be directly measured with AMS, given a similar matrix as that of the iron meteorite. Because this detection limit is still higher by three to four orders of magnitude than the concentrations typical of terrestrial samples, chemical procedures for the extraction or concentration of osmium are needed for the application of AMS to terrestrial samples.

The specific characteristics of AMS for the measurement of osmium, the detection limit of ~20 p.p.b. and the absence of isobaric interference at mass 187, makes it possible to replace the complicated extraction and separation procedures necessary in other methods^{2,11,17} with a simple concentration technique. We used a nickel sulphide fire-assay method developed for the concentration of platinum group elements (PGE)¹⁸ and later modified for the detection of trace amounts of PGE^{19,20}. In this method, the sample (up to 25 g) is mixed with nickel powder, sulphur, and flux material and then fused at temperatures of 1,000 °C. On cooling, nickel sulphide together with the PGE sulphides separate out into a button in the silicate glass. The button is then dissolved in HCl which leaves a residue behind consisting mainly of PGE sulphides and some nickel sulphide. The whole two-step method takes ~24 h and increases the osmium concentration by about four orders of magnitude²¹. It recovers >90% of the PGE metals, including osmium²².

We measured three types of samples with this method (Table 2): two samples from the East Clearwater crater in Quebec, Canada; four samples from the Ries crater in Germany; and a PGE standard (SARM-7) with 63 p.p.b. of osmium²³ which was diluted with silica powder to determine the detection limit and the osmium contribution from the blank. The two samples from

Table 2 Isotope ratios and counting rates for ^{189}Os

Sample no.	Material	Os (p.p.b.)	Re (p.p.b.)	^{189}Os (c.p.m.)	$^{188}\text{Os}/^{189}\text{Os}$	$^{187}\text{Os}/^{189}\text{Os}$	$^{187}\text{Os}/^{186}\text{Os}^*$
Canyon Diablo							
1	Iron meteorite	2000†	200†	13,940	0.791 ± 0.018	0.101 ± 0.004	1.03 ± 0.04
East Clearwater							
2	DCW-2-63-1060 (melt rock)	55.5‡	3.0‡	55,000	0.81 ± 0.08	0.09 ± 0.01	0.91 ± 0.11
3	DCW-2-63-1110 (melt rock)			77,300	0.83 ± 0.01	0.10 ± 0.01	1.02 ± 0.10
Ries							
4	DRG-8-73-1 (gneiss-weakly shocked)	0.165§	0.280§	222	0.89 ± 0.12	0.67 ± 0.04	6.8 ± 0.4
5	Otting (suevite)			544	0.78 ± 0.07	0.79 ± 0.23	8.0 ± 2.3
6	Polsingen (melt rock)			92	0.80 ± 0.03	0.92 ± 0.05	9.3 ± 0.5
7	DRG-1-73-6 (gneiss-inclusion in suevite)	0.006§	0.132§	114	0.93 ± 0.05	1.413 ± 0.10	14.3 ± 1.0
Sarm7							
8	Sarm7/3 (1/1000)	$\sim 0.063\parallel$		131	0.72 ± 0.14	0.29 ± 0.06	2.9 ± 0.6
9	Sarm7/4 (1/10,000)	$\sim 0.0063\parallel$		74	0.63 ± 0.08	0.92 ± 0.05	9.3 ± 0.5
10	Blank	$\sim 0.007¶$		15	0.70 ± 0.07	1.62 ± 0.29	16.4 ± 3.0

* Calculated; † from ref. 1; ‡ from ref. 24; § from ref. 7; || from ref. 23; ¶ corresponds to <0.2 ng Os in the blank.

East Clearwater come from the subsurface melt layer found in this crater. The high osmium concentration (55.5 p.p.b.) measured for one of these samples has been part of the evidence for the extraterrestrial origin of this crater²⁴. Counting rates from untreated Clearwater samples were too low to allow reliable measurements, probably because of the silicate matrix of these samples which, due to its low thermal conductivity and the resulting heat buildup, prevents the effective formation of negative ions. After preconcentration, the counting rates found for these samples were ~ 5 times higher than those of the Canyon Diablo. These counting rates and the associated osmium concentrations show that osmium ionization from the preconcentrated samples is less efficient than that observed for the iron meteorite, probably related to the high electron affinity of the sulfide matrix in the treated samples. The previously unmeasured isotopic ratios of the two Clearwater samples agree with those of Canyon Diablo, a result which strongly supports the extraterrestrial origin of the osmium in this melt layer.

The final set of samples comes from the Ries crater in southern Germany. Although the formation of this crater is, in general, also attributed to the impact of a meteorite^{25,26}, a suite of samples from this crater investigated for traces of extraterrestrial material on the basis of element distributions displayed essentially crustal composition⁷. We measured two samples from this suite with known rhenium and osmium concentrations and two Ries samples with unknown concentrations. Because of the very low concentrations and the relatively high Re/Os ratios, corrections for contributions from the blank and from ^{187}Re have to be considered in the Ries samples. Correction for ^{187}Re was calculated from the counting rate for mass 185 which was around 10 c.p.m. in these samples as compared with rates between 150 and 400 c.p.m. for mass 187. If the blank contribution is taken into account for the two samples with known osmium concentrations, $^{187}\text{Os}/^{186}\text{Os}$ ratios change to 6.4 and 12.0 for samples four and seven, respectively.

Even given these corrections, the Ries samples have a distinctly different isotopic signature from that of the Canyon Diablo meteorite: ^{187}Os is significantly enriched in all the

Ries samples compared with the meteoritic samples, demonstrating a predominantly crustal character. The isotopic ratios are, however, at the low end of the spectrum predicted for crustal values. In view of the few samples measured and the preliminary nature of these measurements, it is too early to draw any firm conclusions with regard to the presence of extraterrestrial material in the Ries crater.

The main point for measuring these samples was to test the method. A measure of the efficiency of the concentration method and ionization can be gained by comparing counting rates for samples with different osmium concentrations, all of which have undergone the same chemical treatment (Fig. 2). The observed correlation suggests that the efficiencies of concentration method and of ion formation are independent of the level of osmium in the range investigated here. Counting rates may perhaps also be used for a rough estimate of osmium concentrations, although isotope dilution methods would obviously give more accurate results.

The present detection limit of the method lies at ~ 0.01 p.p.b. assuming that a counting rate higher than 100 c.p.m. is necessary to obtain high-quality results. These experiments were the first attempt to measure osmium with this method and several factors (such as ionization efficiency, and transmission in the accelerator) have not yet been optimized. It seems realistic to expect that an optimization of the instrument parameters and a concurrent lowering of the osmium contribution from the blank (nickel powder) could result in a further lowering of the detection limit by one order of magnitude. We are also confident that optimization of the method will improve the precision from the present level ($\sim 10\%$, 1σ) to 5% routinely and to 1 or 2% where counting statistics are sufficient.

The results demonstrate that the application of accelerator mass spectrometry after preconcentration of osmium allows the determination of the isotopic composition of osmium in geological samples with osmium levels as low as 0.01 p.p.b. (corresponding to 0.2 ng in the sample), a level significantly below the detection limits reported for other methods²⁻⁴. Because rhenium is found to form negative ions at a much lower rate than osmium,

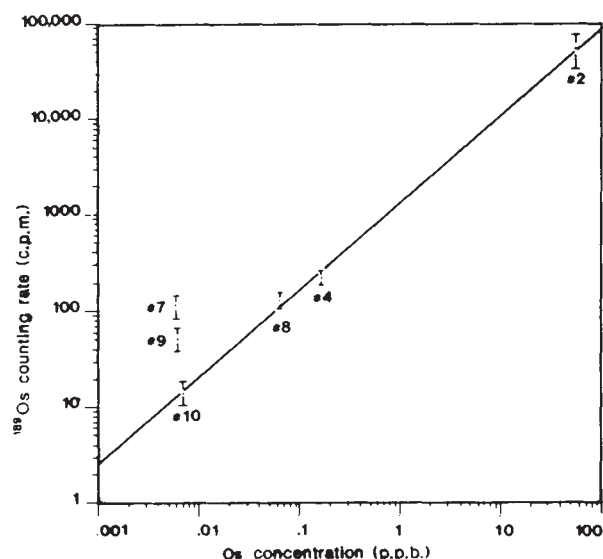


Fig. 2 Counting rates for ^{189}Os as function of osmium concentration in the samples before preconcentration. Samples 7 and 9 were not included in the calculation of the regression line because of the blank contribution to their counting rate. The regression line has a slope of 0.91 with a variance of 0.013 suggesting a linear correlation between these two parameters.

the method does not require the separation of rhenium from osmium but only a relatively straightforward two-step procedure for the preconcentration of osmium. The method lowers the detection limit to a level typical for crustal rocks so that it can be applied without major modifications to most geological problems amenable to the rhenium/osmium method.

We thank K. Nishizumi, F. Hörz, M. R. Dence, R. A. F. Grieve, F. Sellschop and M. Tredoux for providing samples, M. Tredoux, J. C. Van Loon, E. L. Hoffman, P. Wee and S. Kallman for suggestions on the chemical method, and L. Tubbs and T. Hemmick for help during the experiments. This research was supported in part by a grant from the NSF.

Received 12 May; accepted 1 August 1986.

- Allège, C. J. & Luck, J. M. *Earth planet Sci. Lett.* **48**, 148-154 (1980).
- Russ, G. P., Bazan, J. M. & Date, A. R. *Nucl. Chem. Div. A Rep. FY84* (Lawrence Livermore National Laboratory, Livermore, 1984).
- Fassett, J. D., Moore, L. J., Travis, J. C. & DeVoe, J. R. *Science* **230**, 262-267 (1985).
- Palmer, M. R. & Turekian, K. K. *Nature* **319**, 216-220 (1986).
- Luck, J. M. & Turekian, K. K. *Science* **222**, 613-615 (1983).
- Morgan, J. W. & Lovering, J. F. *Earth planet Sci. Lett.* **3**, 219-224 (1967).
- Morgan, J. W., Janssens, M.-J., Hertogen, J., Gros, J. & Takahashi, H. *Geochim. cosmochim. Acta* **43**, 803-815 (1979).
- Turekian, K. K. *Geol. Soc. Am. Spec. Pap.* **190**, 243-247 (1982).
- Turekian, K. K. & Luck, J. M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 8032-8034 (1984).
- Luck, J. M. & Allège, C. J. *Nature* **302**, 130-132 (1983).
- Lindner, D. A. *et al. Nature* **320**, 24-248 (1986).
- Faure, G. *Principles of Isotope Geology* (Wiley, New York, 1977).
- Luck, J. M. & Allège, C. J. *Earth planet Sci. Lett.* **61**, 291-296 (1982).
- Nier, A. O. *Phys. Rev.* **52**, 885 (1937).
- Elmore, D., Conard, N., Kubik, P. W. & Fabryka-Martin, J. *Nucl. Instrum. Meth.* **B5**, 233-237 (1984).
- Elmore, D. *et al.* **286**, 138-140 (1980).
- Luck, J. M. thesis, Univ. Paris 7 (1982).
- Robert, R. V. D., van Wyk, E. & Palmer, R. *Nat. Inst. Metall. Johannesburg, S. Af. Rep.* **1371**, 1-11 (1971).
- Annergarn, H. J., Erasmus, C. S., Sellschop, J. P. F. & Tredoux, M. *Nucl. Instrum. Meth.* **218**, 33-38 (1983).
- Annergarn, H. J., Erasmus, C. S. & Sellschop, J. P. F. *Nucl. Instrum. Meth.* **B3**, 181-184 (1984).
- Teng, R. T. D. thesis, Univ. Rochester (1986).
- Hoffman, E. L., Naldrett, A. J. & Van Loon, J. C. *Analyt. chem. Acta* **102**, 157-166 (1978).
- Sellschop, J. P. F. & Tredoux, M. Preprint, Univ. Witwatersrand, Johannesburg, Council Sci. Ind. Res.: Schonland Res. Cen. Nucl. Sci. (1985).
- Palme, H., Göbel, E. & Grieve, R. A. F. *Proc. 10th Lunar planet. Sci. Conf.*, 2465-2492 (1979).
- Werner, E. *Alb. Bl. Schwab. Albverins* **16**, 153-167 (1906).
- Shoemaker, E. M. & Chao, E. C. T. *J. geophys. Res.* **66**, 3371-3378 (1961).

K-Ar isochron dating of Zaire cubic diamonds

S. Zashu*, M. Ozima* & O. Nitoh†

* Geophysical Institute, University of Tokyo, Tokyo 113, Japan

† Faculty of Engineering, Tokyo University of Agriculture and Technology, Tokyo 184, Japan

U-Pb, Rb-Sr (ref. 1) and Sm-Nd systematics² in inclusions in diamonds suggest strongly that some diamonds are at least 2.0 Gyr old. Such antiquity is also inferred for some diamonds from their very primitive helium isotopic composition³. However, there has been almost no direct radiometric dating of diamonds, except for conventional K-Ar dating^{4,5}, the results of which are questionable due to the possible presence of excess ^{40}Ar . To avoid this problem, we have applied a K-Ar isochron dating method⁶ to ten diamonds from Zaire. The experimental data show good linear correlations in both the ^{40}Ar -K and $^{40}\text{Ar}/^{36}\text{Ar}$ -K/ ^{36}Ar diagrams. These correlations must reflect either an isochron-type relationship or the trapping of an unknown component in the diamonds. In the former case the anomalously high age (6.0 Gyr) casts doubt on the assumption of uniformity in the isotopic abundance of ^{40}K .

The diamonds studied are all from Zaire, and are commonly described as cuboid. The stones are nearly identical, with a nearly perfect cubic shape, a similar size (with edges of ~3 mm and weighing ~0.2 g), and are opaque and slightly yellowish in colour. Microscopic observation reveals that the diamonds are aggregates of numerous micro-crystals of diamond with slightly differing crystal orientations. We observed no trace of impurities or inclusions on the crushed pieces (<100 μm) under a microscope. Although it is impossible to prove, the uniformity of morphology and provenance (Zaire) suggest that they are from a single geological formation. However, the latter point should be judged from the results on an isochron plot.

Experiments were carried out in two phases. The first five stones were analysed for potassium by INAA (instrumental neutron activation analysis) and for argon by mass spectrometry. Two years later, the remaining five stones were analysed with an entirely new mass spectrometer and gas extraction-purification system, and with an INAA technique slightly modified from that of the first experiment. Detailed procedures for Ar and K analyses are given elsewhere⁴. For INAA (integrated neutron flux $\approx 10^{16}$ neutrons per cm^2 with a fast to thermal ratio of 1/3), three standards were used (KCl solution adsorbed on filter paper), with K concentrations spanning the range of diamond K contents. Calibration to the standard samples gave excellent linearity. After the K analyses, the samples were analysed for Ar isotopic contents.

The experimental results are plotted in both ^{40}Ar -K (Fig. 1a) and $^{40}\text{Ar}/^{36}\text{Ar}$ -K/ ^{36}Ar (Fig. 1b) diagrams. Analytical data are given in Table 1. Both plots show good linear correlations, which are most easily understood as isochrons; that is, the observed linear arrays reflect the correlation between K and K-derived ^{40}Ar in the diamonds. The formal age calculated from the slopes of these plots is 6.0 ± 0.3 Gyr; however, this is older than the Earth and is therefore unacceptable. Nevertheless, we believe that the extraordinary age cannot be attributed to mistakes in experimental procedure, as the experiments were carried out in two independent phases with different instruments and techniques, both of which gave very similar results (see Fig. 1). There are several possible explanations for this extraordinary result.

First, the observed linear correlation may have nothing to do with an isochron. The linear correlation in Fig. 1a may then be explained by assuming the incorporation of variable amounts of an unknown component which had a constant $^{40}\text{Ar}/\text{K}$ ratio. This component must have contained excess ^{40}Ar relative to in

Table 1 K, ^{40}Ar and ^{36}Ar contents in diamonds

Sample no. (Weight (g))	K (p.p.m.)	Hot blank (2,050 °C)		Diamonds			
		^{36}Ar ($\times 10^{-10}$)	^{40}Ar ($\times 10^{-7}$)	^{36}Ar ($\times 10^{-10}$)	^{40}Ar ($\times 10^{-7}$)	K/ ^{36}Ar (p.p.m./ 10^{-10})	$^{40}\text{Ar}/^{36}\text{Ar}$
831101 (0.22594)	12.3 $\pm$ 0.4	0.54	0.12	1.28 $\pm$ 0.28	20.4 $\pm$ 0.2	9.60 $\pm$ 2.40	15,900 $\pm$ 3,600
831102 (0.20048)	47.1 $\pm$ 1.4	0.86	0.21	3.71 $\pm$ 0.45	87.2 $\pm$ 0.6	12.71 $\pm$ 1.91	23,500 $\pm$ 4,300
831103 (0.13558)	18.5 $\pm$ 0.5	0.76	0.17	3.88 $\pm$ 0.42	36.1 $\pm$ 0.1	4.75 $\pm$ 0.65	9,300 $\pm$ 1,000
831104 (0.13412)	22.7 $\pm$ 0.9	0.86	0.17	1.51 $\pm$ 0.44	44.0 $\pm$ 0.1	15.04 $\pm$ 4.98	29,100 $\pm$ 8,600
831105 (0.13349)	14.2 $\pm$ 0.4	0.88	0.19	1.30 $\pm$ 0.45	26.0 $\pm$ 0.2	10.96 $\pm$ 4.04	20,400 $\pm$ 7,200
850702 (0.18614)	27.8 $\pm$ 1.0	1.06	0.14	2.43 $\pm$ 0.43	48.0 $\pm$ 0.11	11.44 $\pm$ 2.44	19,700 $\pm$ 3,500
850703 (0.23855)	14.8 $\pm$ 0.5	0.50	0.13	23.5 $\pm$ 0.3	25.8 $\pm$ 0.09	0.63 $\pm$ 0.03	1,100 $\pm$ 20
850704 (0.19140)	10.2 $\pm$ 0.4	1.09	0.22	2.42 $\pm$ 0.49	18.3 $\pm$ 0.12	4.21 $\pm$ 1.02	7,600 $\pm$ 1,600
850705 (0.21891)	22.9 $\pm$ 0.8	0.86	0.13	3.08 $\pm$ 0.44	34.1 $\pm$ 0.09	7.44 $\pm$ 1.32	11,100 $\pm$ 1,600
850707 (0.21712)	29.1 $\pm$ 1.0	0.93	0.16	2.24 $\pm$ 0.40	52.4 $\pm$ 0.12	12.99 $\pm$ 2.77	23,400 $\pm$ 4,200

Argon contents (sample and blank) are in cm^3 STP per g. Errors (1σ) in K are statistical errors; errors (1σ) in Ar in diamonds include analytical error in mass spectrometry and 50% error in the hot blank.

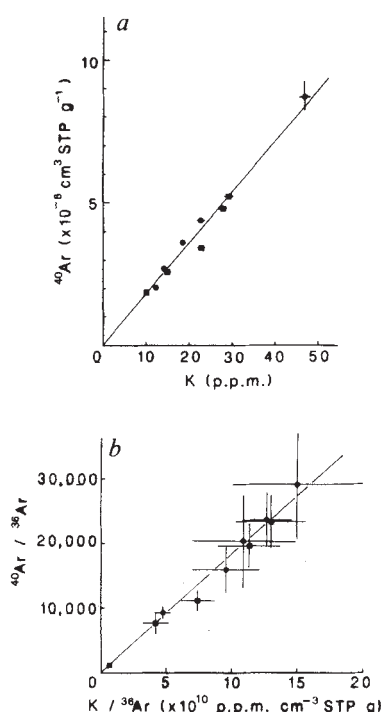


Fig. 1 ^{40}Ar -K (a) and $^{40}\text{Ar}/^{36}\text{Ar}$ -K/ ^{36}Ar (b) diagrams. Circles, experiments carried out in 1983; squares, experiments carried out in 1986 with a new mass spectrometer and sample extraction line. Larger errors in b than in a are due to ^{36}Ar hot blank corrections. Intercepts with the vertical axes are $^{40}\text{Ar} = (-0.14 \pm 0.23) \times 10^{-6} \text{ cm}^3 \text{ STP g}^{-1}$ (a), $^{40}\text{Ar}/^{36}\text{Ar} = -700 \pm 900$ (b).

situ ^{40}K , to give the extraordinarily old age. Such constraints seem to require that the unknown component was in a fluid state, to ensure the homogenization necessary to give a constant $^{40}\text{Ar}/\text{K}$ ratio, and to acquire excess ^{40}Ar from outside the system. If the linear correlation were due to incorporation of such a component, however, we would not see a linear array in the $^{40}\text{Ar}/^{36}\text{Ar}$ -K/ ^{36}Ar diagram (Fig. 1b). This linear array could reflect errors in the measurement of ^{36}Ar for each diamond, such as could result from the incorporation of variable amounts of air argon or from inadequate correction for hot blank ^{36}Ar . However, to generate the observed large spread in Fig. 1b would require several times as much ^{36}Ar as was observed in the hot blank, which we consider to be rather unlikely.

Alternatively, if we accept the linear correlation as an isochron, we are forced to question the constancy of the decay constant or the isotopic composition of potassium. The radioactive decay of ^{40}K to ^{40}Ar is governed by an innermost orbital electron, the behaviour of which may be affected by environmental conditions. High hydrostatic pressure in the deep mantle would enhance the decay rate by compressing the electronic shell. However, an experimental investigation of the electron-capture decay of ^7Be (oxide) at a hydrostatic pressure of ~ 300 kbar yielded an increase in decay rate of only 0.55% (ref. 7).

A more plausible speculation postulates that the diamonds may differ in potassium isotopic composition from common potassium (present $^{40}\text{K}/\text{K} = 1.17 \times 10^{-4}$). Metallic grains participating in the Earth's accretion were subjected to cosmic-ray irradiation, and must have contained a significant amount of spallogenic ^{40}K , as is seen in iron meteorites⁸. As metallic iron cannot easily accommodate substantial amounts of potassium, the potassium in such cosmic-ray-irradiated metallic grains must have been essentially pure cosmogenic ^{40}K . The metallic grains were then incorporated into planetesimals, and accreted onto

the Earth⁹. If core formation proceeded rapidly^{10,11}, a substantial proportion of the cosmogenic ⁴⁰K could be carried into the core without reaching isotopic equilibrium with surrounding silicates. The base of the mantle could thus be considerably enriched in cosmogenic ⁴⁰K. (A similar mechanism was invoked to explain solar-type helium in diamonds¹².) It is therefore conceivable that some diamonds trapped such ⁴⁰K-enriched potassium. If the ⁴⁰K/K ratio in the diamonds were twice as large as that in common potassium, the age would be reduced from 6.0 to 4.6 Gyr.

Obviously none of the above interpretations is conclusive. However, it is clear that the observed correlations in the ⁴⁰Ar-K and ⁴⁰Ar/³⁶Ar-K/³⁶Ar diagrams are not artefacts, but reflect some unknown process(es) occurring in the mantle, or during the Earth's accretion.

We thank Drs T. Kirsten and I. P. Wright for stimulating comments and Dr. M. H. Dodson for a constructive review. Neutron irradiations were performed at Rikkyo University, for which we thank Professor K. Tomura.

Received 30 April; accepted 28 July 1986.

1. Kramers, J. D. *Earth planet. Sci. Lett.* **42**, 58-70 (1979).
2. Richardson, S. H., Gurney, J. J., Erlank, A. J. & Harris, J. W. *Nature* **310**, 198-202 (1984).
3. Ozima, M. & Zashu, S. *Science* **219**, 1067-1068 (1983).
4. Ozima, M., Zashu, S. & Nitoh, O. *Geochim. cosmochim. Acta* **47**, 2217-2224 (1983).
5. Ozima, M., Takaoka, N., Nitoh, O. & Zashu, S. in *Materials Science of the Earth's Interior* (ed. Sunagawa, I.) 375-386 (Terra Scientific, Tokyo, 1984).
6. Hayatsu, A. & Carmichael, C. M. *Can. J. Earth Sci.* **14**, 337-345 (1977).
7. Hensley, W. K., Bassett, W. A. & Huizenga, J. R. *Science* **181**, 1164-1165 (1973).
8. Imamura, M., Shima, M. & Honda, M. *Z. Naturf.* **35A**, 267-279 (1980).
9. Hayashi, C., Nakazawa, K. & Nakagawa, Y. in *Protostars and Planets Vol. 2* (eds Black, D. C. & Matthews, M. S.) 1100-1154 (University of Arizona Press, Tucson, 1985).
10. Stevenson, D. J. *Science* **214**, 611-619 (1981).
11. Sasaki, S. & Nakazawa, K. *J. geophys. Res.* (in the press).
12. Becker, R. H. & Pepin, R. O. *Earth planet. Sci. Lett.* **70**, 1-12 (1984).

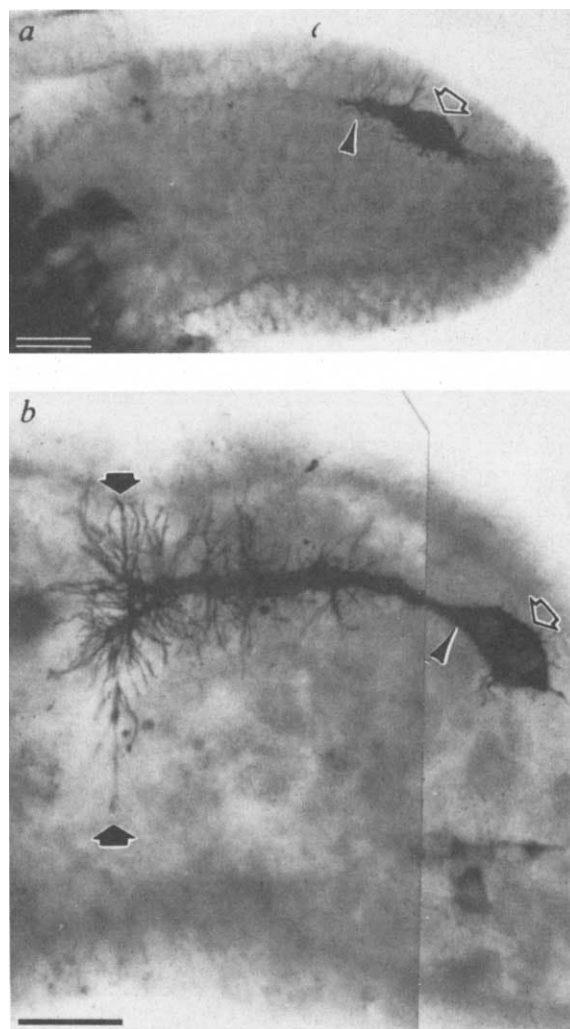


Fig. 1 Pathfinding by pioneer Ti1 neurones in a control limb bud in culture. *a*, Pioneer neurone pair (open arrow) undergoing axonogenesis at the limb tip in a normal (not cultured) 31-32% stage embryo. This is the stage at which experiments were begun. The arrowhead indicates the leading edge of the growth cone. *b*, Pioneer neurones (open arrow) in a different embryo after 30 h in culture. The growth cones have grown proximally along the limb axis and have reached the trochanter/coxa limb segment boundary, where they have extended long, circumferential filopodia (solid arrows). Numerous filopodia are extended ahead of the growth cones and laterally from the axons. The arrowhead indicates the position of the growth cones at the start of the culture period, estimated from the positions of growth cones in embryos from the same egg clutch which were fixed at the onset of the culture period. Metathoracic limb buds (dorsal, up; proximal, left). Calibration bars, 50 μ m. Culture medium (modified for grasshopper embryos by F. Lefcort): RPMI-1640; 0.2% sodium carbonate; 1 μ M oxaloacetic acid; 0.45 mM sodium pyruvate; 2 mM glutamine; 0.45% D(+)-glucose; 0.2 IU ml⁻¹ insulin; 10⁻⁶ β -ecdysterone; 50 μ g ml⁻¹ gentamycin.

Disoriented pathfinding by pioneer neurone growth cones deprived of filopodia by cytochalasin treatment

David Bentley & Alma Toroian-Raymond

Department of Zoology, University of California, Berkeley, California 94720, USA

A major question in developmental neurobiology is how developing nerve cells accurately extend processes to establish connections with their target cells. This problem involves both the nature of cues for growth cone guidance and also the question of how growth cones survey their environment for cues and respond by altering their direction of migration. The filopodia which normally extend from neuronal growth cones have been shown to affect growth cone steering *in vitro*¹⁻⁴ and it has been proposed that they function *in vivo* in the detection of and response to guidance cues⁵⁻¹⁰. This hypothesis could be tested *in vivo* if growth cones which normally have filopodia could be induced to migrate in their absence. The pair of Ti1 neurones are the first neurones to extend axons through the limb buds of embryonic grasshoppers¹¹. We report here an examination of the migration of Ti1 pioneer growth cones deprived of filopodia by culture in agents which disrupt actin microfilaments. Under these conditions, axons continue to extend but a large percentage of growth cones are highly disoriented. Our results indicate that Ti1 filopodia are not necessary for axonal elongation *in vivo* but that they are important for correctly oriented growth cone steering.

Embryos of *Schistocerca americana* were dissected from eggs at the 31-32% stages of development, when the Ti1 pioneer neurones are undergoing axonogenesis, and divided into three groups. One group was fixed immediately to establish the

average initial position of the pioneer growth cones (Fig. 1*a*; growth cone morphology defined as in refs 12 and 13); a second group was maintained in normal medium for 24-36 h before fixation to serve as a culture control (Fig. 1*b*); the third group was cultured in the presence of filopodium-suppressing agents. The dorsal closures of embryos were opened to expose their interiors to externally applied solutions. For each experiment, embryos from a single egg clutch were used.

Exposure of growth cones to the actin microfilament-disrupting agent cytochalasin B prevents the extension of

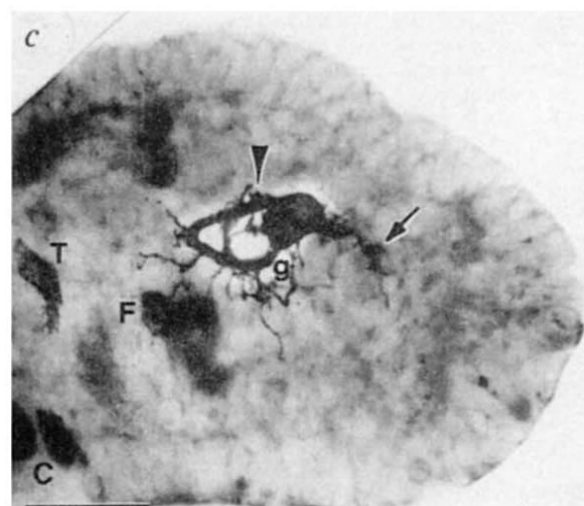
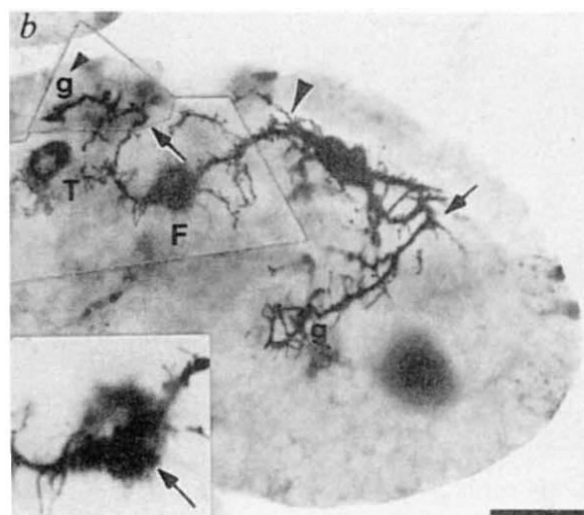
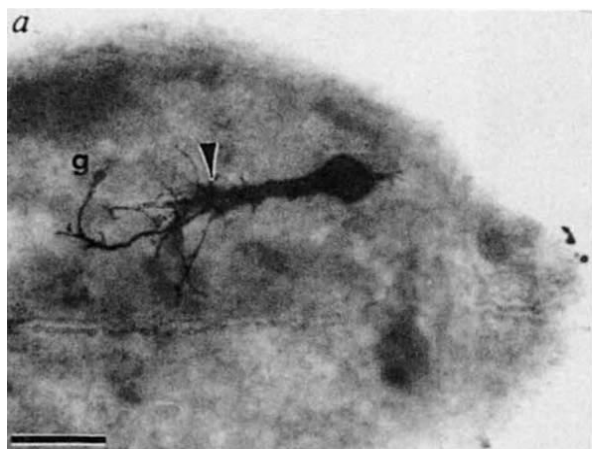


Fig. 2 Disoriented pathfinding by pioneer growth cones deprived of filopodia. *a*, Pioneer neurones after 24 h in culture with $1.0 \mu\text{g ml}^{-1}$ cytochalasin B. The growth cone (g) is devoid of filopodia. The axon has turned until it is heading slightly distally (to right) and is more than 90° off the normal proximal course. *b*, Pioneer neurones after 36 h in culture with $0.1 \mu\text{g ml}^{-1}$ cytochalasin B. From the cell bodies, axons extend both proximally (normal direction) and distally (abnormal). Growth cones (g, g') of both axons are devoid of filopodia. Apparently, one axon (g') extended proximally from the cell body and contacted guidepost neurone Fe1 (F). Leaving this neurone, it curved almost 180° dorsally and distally (left arrow), and finally turned approximately 90° in the proximal direction. The other axon (g) extended distally, made a 90° ventral turn (right arrow) along the limb circumference and finally turned more than 180° to intersect itself (g, right). The inset (lower left) shows the pioneer axon (arrow) spreading extensively across the surface of neurone Fe1. T, guidepost neurone Tr1. *c*, Pioneer neurones after 36 h in culture with $0.1 \mu\text{g ml}^{-1}$ cytochalasin B. The axons extended proximally (to left) and then turned 180° to contact the cell bodies (growth cones, g). A major process is extended distally (arrow). F, guidepost neurone Fe1; T, guidepost neurone Tr1; C, guidepost neurone pair Cx1. Arrowheads indicate the inferred position of the growth cone in *a*, *b* and *c* at the onset of the culture period. Metathoracic limbs (dorsal, up; proximal, left). Calibration bars, $50 \mu\text{m}$.

filopodia *in vitro*^{14,15}. We used cytochalasin B in most experiments but all results were confirmed in dihydrocytochalasin B¹⁶ and in cytochalasin D, which was effective at lower concentrations (Table 1). Embryos were cultured in a series of concentrations ranging from 0.05 to $10 \mu\text{g ml}^{-1}$. To evaluate the onset and duration of drug effects, embryos were cultured for 0.5, 1, 3, 8 and 24 h in cytochalasin B. *In vitro*, most filopodia were suppressed within minutes of exposure to cytochalasin B¹⁵; *in vivo*, filopodia were suppressed within 0.5 h and remained suppressed for at least 8 h. Experimental culture solutions were changed every 8 h since a few filopodia reappear by 24 h. In 75 experimental embryos, the routes of 500 T11 pioneer neurones were examined; routes in 38 embryos are quantified in Table 1. Neurones were visualized by indirect labelling of fixed tissue with anti-horseradish peroxidase (HRP) antibody, which selectively binds to insect neurones^{13,17}.

In cultured embryos, T11 pioneer growth cones developed the normal array of filopodia and growth along the initial proximal portion of their route was completed (Fig. 1*b*). No instances of deviation from the normal axial route were observed (Table 1).

Culturing in the presence of agents which disrupt actin microfilaments suppressed extension of filopodia. Growth cones appeared as enlargements at the ends of axons, but extended neither filopodia nor lamellipodia (Figs 2, 3). Along the axon behind the growth cone, thin processes were sometimes observed which differed from normal filopodia in that they were thicker, were not of uniform diameter, and had frequent sharp turns¹⁴.

Several features demonstrate that pioneer axons are able to elongate even when their growth cones are devoid of filopodia: (1) experimental axons extended well beyond the growth cones of control embryos fixed at the onset of the experiment (Figs 2, 3); (2) experimental axons often had extensive segments (some $>100 \mu\text{m}$) of abnormally oriented growth (Fig. 2*b*); (3) axons, or major branches extending distally from the cell body, were regularly observed (Fig. 2*b, c*): these are extremely rare in normal embryos¹⁸.

In vitro, axons with growth cones deprived of filopodia are unable to elongate on substrates of low adhesivity¹⁵ but do elongate on substrates of high adhesivity¹⁴. The ability of pioneer axons to elongate *in vivo* in the absence of filopodia therefore suggests that the substrate *in vivo* has relatively high adhesivity.

Many growth cones deprived of filopodia migrated along abnormal routes. They displayed a variety of turns, some growing circumferentially, some distally and some in circles (Fig. 2). We quantified the extent of disorientation according to the degree of deviation from proximal axial growth (Table 1). Since many axons meandered through the limb along a route which included several turns, we scored both the overall direction of the axon and also the most disoriented short segment (at least $20 \mu\text{m}$). Under the most disruptive conditions tested ($0.1 \mu\text{g ml}^{-1}$ cytochalasin D; Table 1) 44% of axons were disoriented in their overall direction and 84% had disoriented segments. Overall, 30% of the 247 growth cones scored has disoriented axons and 69% had disoriented segments of growth (Table 1).

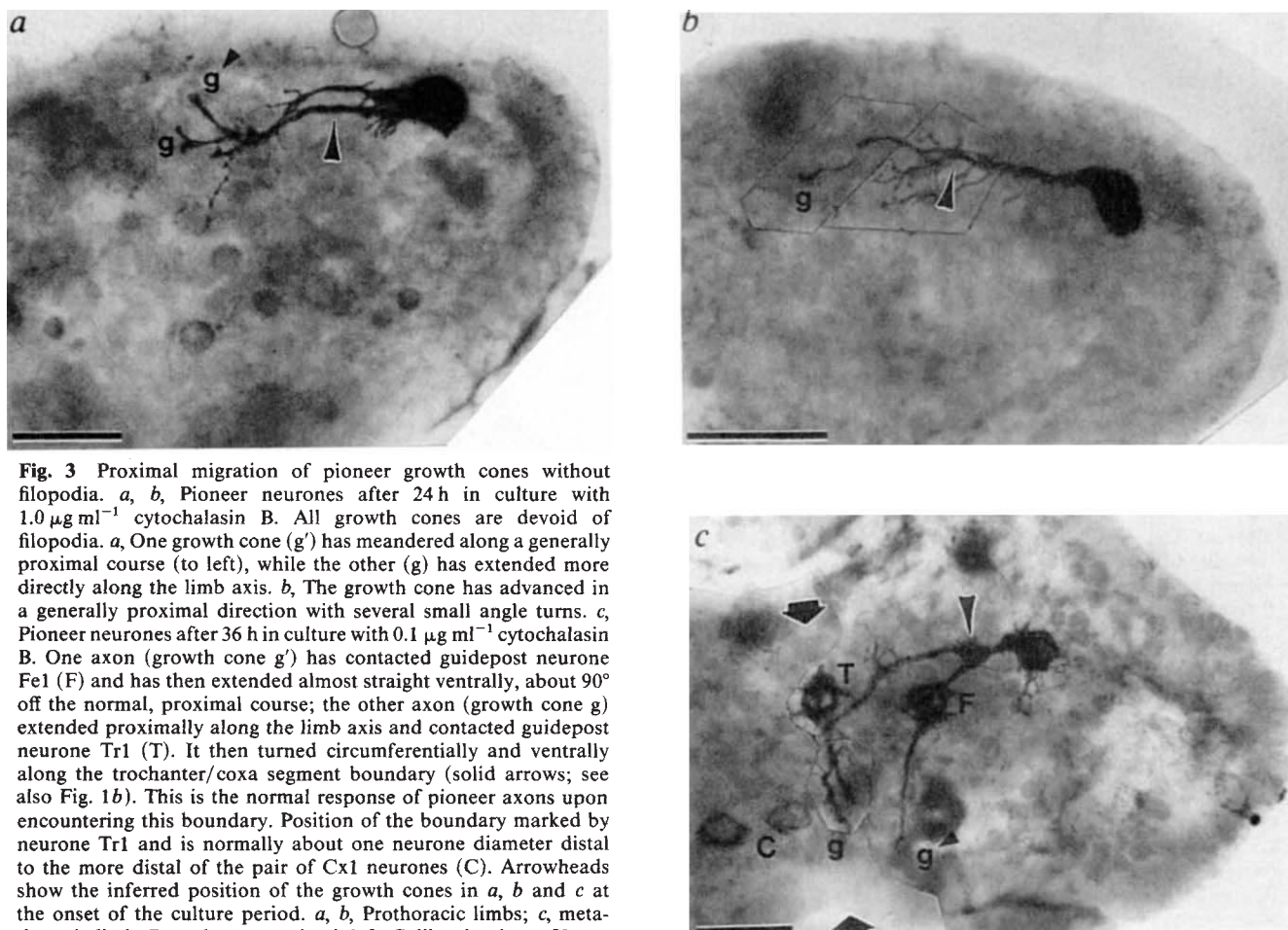


Fig. 3 Proximal migration of pioneer growth cones without filopodia. *a, b*, Pioneer neurones after 24 h in culture with $1.0 \mu\text{g ml}^{-1}$ cytochalasin B. All growth cones are devoid of filopodia. *a*, One growth cone (*g'*) has meandered along a generally proximal course (to left), while the other (*g*) has extended more directly along the limb axis. *b*, The growth cone has advanced in a generally proximal direction with several small angle turns. *c*, Pioneer neurones after 36 h in culture with $0.1 \mu\text{g ml}^{-1}$ cytochalasin B. One axon (growth cone *g'*) has contacted guidepost neurone Fe1 (*F*) and has then extended almost straight ventrally, about 90° off the normal, proximal course; the other axon (growth cone *g*) extended proximally along the limb axis and contacted guidepost neurone Tr1 (*T*). It then turned circumferentially and ventrally along the trochanter/coxa segment boundary (solid arrows; see also Fig. 1*b*). This is the normal response of pioneer axons upon encountering this boundary. Position of the boundary marked by neurone Tr1 and is normally about one neurone diameter distal to the more distal of the pair of Cx1 neurones (*C*). Arrowheads show the inferred position of the growth cones in *a, b* and *c* at the onset of the culture period. *a, b*, Prothoracic limbs; *c*, meta-thoracic limb. Dorsal, up; proximal, left. Calibration bars, $50 \mu\text{m}$.

Although many growth cones were disoriented, others continued to extend along a trajectory which approximated the normal route; some turned two or more times but still maintained a generally proximal course (Fig. 3). At least three possibilities could contribute to this: (1) the direction of growth is random, so that approximately 25% growth cones would be expected to orient within the 45° quadrant of proximal growth; (2) growth cones tend to continue in the direction in which they are growing¹⁹ and the growth cones had initiated proximal growth before the filopodia were suppressed; (3) growth cones without filopodia still respond to some degree to the normal guidance cues. The high percentage (56–77%; Table 1) of growth cones which continued to extend proximally seems to rule out a solely random effect and the maintenance of proximal growth through several turns appears to favour the possibility that the growth cones retained a diminished pathfinding capability.

Although growth cone disorientation could be the result of the loss of pathfinding capabilities normally conferred by the presence of filopodia, it could also be due to an effect of cytochalasin treatment on either the expression of guidance cues or on the ability of growth cone membrane to detect or to respond to such cues. In this case, growth cones should fail to respond to guidance cues which they happen to encounter. Two cues to which growth cones normally respond are immature 'guidepost' neurones and the prospective trochanter/coxa limb segment boundary^{8,12,13,18,20–22}. Guidepost neurones are observed by anti-HRP binding in limbs which were allowed to differentiate for 36 h in culture (Figs 2*b, c, 3c*). Normally, pioneer neurones increase membrane apposition to guidepost neurones by spreading filopodia, lamellipodia and branches on their surfaces. In the presence of cytochalasins, pioneer axons

maintain contact with guidepost neurones (Figs 2*b, 3c*) and spread lamellae on their surfaces (Fig. 2*b*, inset) in limbs where the growth cones are disoriented. This suggests that the guidepost neurones continue to express factor(s) recognized by pioneer axons and that the axons remain capable of responding.

Table 1 Frequency of disoriented pathfinding by growth cones devoid of filopodia

Treatment	No. of neurones (embryos)	% Axons $> 45^\circ$	% Segments $> 45^\circ$	% Segments $> 90^\circ$
Control	144 (36)	0	0	0
High	32 (6)	44	84	53
0.1 $\mu\text{g ml}^{-1}$ CD				
Intermediate	90 (12)	36	72	28
(1.0, 0.25 $\mu\text{g ml}^{-1}$ DHCB				
Low	125 (20)	23	62	17
0.1 $\mu\text{g ml}^{-1}$ CB				
0.05 $\mu\text{g ml}^{-1}$ CD				
0.1 $\mu\text{g ml}^{-1}$ DHCB				
Total	247 (38)	30	69	26

Processes are scored for the degree of deviation from straight proximal axial growth in the distal limb bud: processes deviating by more than 45° are considered disoriented. 'Axons' refers to the general course of the axon and 'segments' to axon segments of at least $20 \mu\text{m}$. The most disoriented segment of each axon was scored. CD, cytochalasin D; DHCB, dihydrocytochalasin B; CB, cytochalasin B. Where multiple agents or dosages were used, embryos were divided equally among the categories. High, intermediate and low refer to the relative efficacy of the treatments listed in promoting growth cone disorientation.

Normally, pioneer axons reorient from axial to circumferential growth upon encountering the prospective trochanter/coxa segment boundary, and grow ventrally along that boundary^{12,13,18}. In the presence of cytochalasin, a pioneer axon can respond normally to that boundary in the same limb in which the other pioneer axon is disoriented (Fig. 3c). This indicates that levels of cytochalasin which result in pioneer disorientation can still permit expression of segment boundary recognition factors and that growth cones which encounter the boundary remain capable of responding normally. These observations do not demonstrate that all guidance cues which normally promote proximal axial growth are expressed and interpreted normally in the presence of cytochalasins, but they do show that cytochalasins do not generally debilitate the expression of or response to normal guidance cues.

The degree to which filopodia and lamellipodia are extended from growth cones varies considerably among the types of neurones which have been examined *in vivo*^{10,12,23-27}. In general, growth cones which are migrating along previously established axons or axon tracts tend to have relatively simple growth cones, whereas growth cones migrating over non-neural substrates or entering complex regions of neuropile extend more processes^{9,10,13,23-26,28}. The same growth cone can assume quite different morphologies depending on the nature of the region through which it is migrating^{9,13,25}. Rapidly moving growth cones generally may be simpler and more lamellar in form²⁹, while growth cones which are moving slowly²⁹ or are at choice points^{10,25,26,28} tend to be more filopodial and extended. These observations suggest that the growth cones which have numerous filopodia *in vivo* may be moving relatively slowly through regions

of relatively high ambiguity in guidance information. Our results indicate that in such cases filopodial exploration has an important role in growth cone guidance.

We thank David Weisblat, W. Zacheus Cande and Manfred Schliwa for suggestions and criticisms of the manuscript. Support was provided by NIH Jacob Javits Award NS09074-16.

Received 10 March; accepted 3 June 1986.

1. Bray, D. *J. Cell Sci.* **37**, 391-410 (1979).
2. Bray, D. & Chapman, K. *J. Neurosci.* **5**, 3204-3213 (1985).
3. Letourneau, P. C. in *Molecular Bases of Neural Development* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 269-294 (Wiley, New York, 1985).
4. Wessells, N. K. & Nottall, R. P. *Expl Cell Res.* **115**, 111-122 (1978).
5. Weiss, P. *Symp. Soc. Study dev. Growth* **5** (suppl.), 163-203 (1941).
6. Mason, C. *Trends Neurosci.* **8**, 304-306 (1985).
7. Bastiani, M. J., Raper, J. A. & Goodman, C. S. *J. Neurosci.* **4**, 2311-2328 (1984).
8. Bentley, D. & Keshishian, H. *Science* **218**, 1082-1088 (1982).
9. Taylor, J. S. H. & Roberts, A. *J. Embryol. exp. Morph.* **75**, 49-66 (1983).
10. Tosney, K. W. & Landmesser, L. T. *J. Neurosci.* **5**, 2345-2358 (1985).
11. Bate, C. M. *Nature* **260**, 54-56 (1976).
12. Caudy, M. & Bentley, D. *J. Neurosci.* **6**, 1781-1795 (1986).
13. Caudy, M. & Bentley, D. *J. Neurosci.* **6**, 364-379 (1986).
14. Marsh, L. & Letourneau, P. C. *J. Cell Biol.* **99**, 2041-2047 (1984).
15. Yamada, K. M., Spooner, B. S. & Wessells, N. K. *J. Cell Biol.* **49**, 614-635 (1971).
16. Bray, D. & Gilbert, D. A. *Rev. Neurosci.* **4**, 505-523 (1981).
17. Jan, L. Y. & Jan, Y. N. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2700-2704 (1982).
18. Bentley, D. & Caudy, M. *Cold Spring Harb. Symp. quant. Biol.* **48**, 573-585 (1983).
19. Katz, M. J. *J. Neurosci.* **5**, 589-595 (1985).
20. Bentley, D. & Caudy, M. *Nature* **304**, 62-64 (1983).
21. Ho, R. K. & Goodman, C. S. *Nature* **297**, 404-406 (1982).
22. Keshishian, H. & Bentley, D. *Dev. Biol.* **96**, 89-124 (1983).
23. Kuwada, J. Y. & Kramer, A. P. *J. Neurosci.* **3**, 2098-2111 (1983).
24. Lopresti, V., Macagno, E. R. & Levinthal, C. *Proc. natn. Acad. Sci. U.S.A.* **70**, 433-437 (1973).
25. Mason, C. *J. Neurosci. Res.* **13**, 55-73 (1985).
26. Raper, J. A., Bastiani, M. J. & Goodman, C. S. *J. Neurosci.* **3**, 20-41 (1983).
27. Roberts, A. & Taylor, J. S. H. *J. Embryol. exp. Morph.* **75**, 31-47 (1983).
28. Rusoff, A. C. *J. Neurosci.* **4**, 1414-1428 (1984).
29. Argiro, V., Bunge, M. B. & Johnson, M. I. *J. Neurosci.* **4**, 3051-3062 (1984).

Calbindin immunoreactivity alternates with cytochrome c-oxidase-rich zones in some layers of the primate visual cortex

Marco R. Celio*, L. Schäfer*, John H. Morrison†, A. W. Norman‡ & F. E. Bloom†

* Institute of Anatomy, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

† Research Institute of Scripps Clinic, 10666 North Torrey Pines Road, La Jolla, California 92037, USA

‡ Department of Biochemistry, University of California at Riverside, Riverside, California 9251, USA

Calcium ions have a pivotal role in many neuronal activities¹, but little is known about their involvement in the cortical processing of visual information². Using immunohistochemical methods, we have now detected a calcium-binding protein, calbindin-D-28K (ref. 3, calbindin), which may confer on certain compartments of cortical area 17 the ability to modulate Ca²⁺ metabolism. Thus, calbindin occurs in the primate striate cortex in a pattern almost complementary to that displaying strong cytochrome c-oxidase activity. From this and other observations, we deduce that the distribution of calbindin-immunoreactive sites corresponds mainly to extra-geniculocortical connections of the primary visual cortex. This implies that the geniculocortical and extra-geniculocortical compartments of area 17 differ in an intracellular system for Ca²⁺ homeostasis.

The basic biochemical function of the vitamin D-induced calcium-binding protein⁴ calbindin is to bind calcium intracellularly and to act as an enzyme activator⁵ or as a buffer or shuttle of this cation⁶. Its physiological function is unknown, but studies in the rat hippocampus have demonstrated an inverse relationship between calbindin concentration and cell excitability⁷.

Using antibodies against chicken intestinal calbindin whose specificity has been characterized previously (ref. 8 and Fig. 1

legend), we found that the physical properties of the visual cortex protein recognized by the calbindin antiserum corresponded to those of purified chicken intestinal calbindin (Fig. 1). Examination of sections of visual cortex from six squirrel monkeys (*Saimiri sciureus*) by standard light microscopic immunocytochemical technique demonstrated that calbindin immunostaining in the monkey visual cortex included both cell and neuropil labelling. Sections cut perpendicular to the cortical surface (Figs 2a, 3a) revealed a stratification of the immunostaining pattern with calbindin antisera. All layers contained immunoreactive sites, but the staining intensity varied with the layers. With the exception of layer IVc, the shifts in immunostaining densities corresponded to the laminar boundaries identified by Nissl or cytochrome oxidase staining of adjacent sections (Figs 2b, 3b).

Layers I and upper II displayed a strongly calbindin-positive neuropil (Figs 2a, 3a, d) while only layer I is known to harbour a thin cytochrome-oxidase-positive band in its middle portion (ref. 9 and Fig. 3c). In layers lower II, III, IVc- α and faintly in V, a mosaic of globular 'lacunae' of diminished calbindin immunoreactivity was consistently observed (Figs 2a, 3d, e, 4a, c). These patches had an approximate diameter of 150 μ m, a centre-to-centre spacing of 450-500 μ m and were in exact register in the various layers. They corresponded strictly to the cytochrome-oxidase-rich 'puffs' in layers II, III and V (Figs 3b, c, 4b). Layer IVa showed strong homogeneously distributed calbindin immunoreactivity which merged with lower layer III (Figs 2a, 3a, d, e, 4c) and a reticulated pattern of cytochrome-oxidase-rich sites (Fig. 3b, c). Layer IVb consistently lacked calbindin (Figs 2a, 3a, d, e) but showed faint cytochrome oxidase staining (Fig. 3b, c). The upper part of layer IVc- α contained calbindin-unreactive lacunae in register with those in layers II-III (Figs 2a, 3d, e) and was homogeneously rich in cytochrome oxidase (Figs 2b, 3b, c). The upper half of layer IVc- β displayed strong calbindin immunoreactivity, while that in the lower half was weak or absent (Figs 2a, 3a, d, e). Cytochrome-c-oxidase activity in layer IVc- β showed a complementary picture (ref. 9 and Fig. 3b, c). The calbindin immunostaining

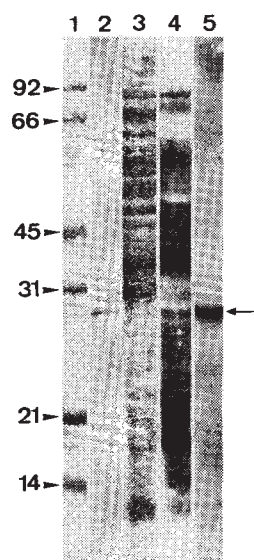


Fig. 1 Immunoblot after SDS-polyacrylamide gel electrophoresis of an extract of visual cortex. The brain of an adult rhesus monkey was rapidly removed and the lateral convexity of the primary visual cortex dissected out and frozen on dry ice. The tissue was homogenized, subjected to electrophoresis and the immunoblot performed as described elsewhere¹⁴. Purified chicken intestinal calbindin (2 μ g, lane 2) and visual cortex extract (7 mg ml⁻¹ protein, lane 3) were run in parallel and blotted on nitrocellulose paper. The visual cortex proteins transferred to the nitrocellulose were visualized with AuroDye (Janssen) and are shown in lane 4. A strip of the same nitrocellulose sheet was incubated with the calbindin antiserum and the bound antibody was revealed using an indirect peroxidase method (lane 5). In the visual cortex, the antiserum used in our study recognized a protein which displayed the same electrophoretic mobility as chicken intestinal calbindin (arrow). In some cases the band defined by the calbindin antiserum was split in two halves. Lane 1, relative molecular mass markers ($\times 10^{-3}$). Lanes 1 and 3 were photographed after the gel had been dried. (The calbindin band is not visible in the Coomassie blue stained gel (lane 3) because of the relative insensitivity of this dye (microgram range) compared with the immunological reactivity (nanogram range, lane 5).)

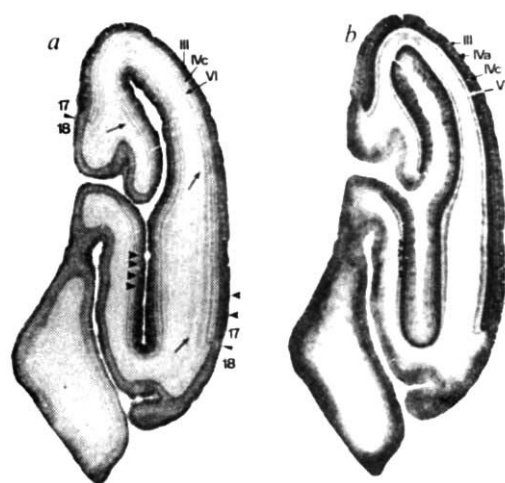


Fig. 2 Low-power pictures of two consecutive parasagittal sections through the squirrel monkey occipital lobe. *a*, Section incubated with calbindin antiserum. A parcellation in two different areas (17 and 18, arrow-heads mark the border) is evident. There are no differences in laminar staining for calbindin between central or peripheral retinal representation. The stratification of the cortex is particularly marked in area 17. The third layer (III) of the primary visual cortex displays a strong but discontinuous staining (arrow-heads indicate calbindin-negative lacunae; compare with *b*). Layer IV is divided into four sublaminae. Note the conspicuous bundle of thin axons in the white matter (arrows). Animal 2/83. $\times 12$. *b*, Consecutive section incubated for cytochrome-*c* oxidase histochemistry (performed according to ref. 19). The puffs in layers II–III (arrow-heads) correspond to the calbindin lacunae of *a*. The high activities in layers IVa, IVc and VI are characteristic. $\times 12$.

Methods. The primary visual cortex (area 17, VI) of six adult squirrel monkeys (*S. sciureus*) of both sexes was studied. The animals were anaesthetized with Nembutal and perfused with cold fixative (4% paraformaldehyde in 0.1 M phosphate buffer pH 7.3). After 15 min perfusion, the brain was dissected out and post-fixed for 12 h in the same fixative. After several washes with 18% sucrose, the tissue was frozen on pulverized dry ice and sectioned in a cryostat. Coronal and parasagittal sections were 50 μ m thick and tangential sections were 20 μ m thick. The free-floating sections were preincubated with 0.4% Triton X-100 in 0.1 M Tris-HCl buffer pH 7.3, then incubated with the antiserum against chicken intestinal calbindin-D-28K diluted 1:5,000 to 1:10,000 in Tris buffer. After incubation on a shaker at 4°C for 72 h, the sections were rinsed in buffer and incubated for 2 h with biotinylated goat anti-rabbit IgG and then with the avidin-peroxidase complex (Vectastain, Vector Laboratories) or by the unlabelled antibody method. The sections were incubated for 10–20 min in the developing solution containing 3,3'-diaminobenzidine-tetrahydrochloride (0.04%) and H₂O₂ (0.004%) to reveal the bound peroxidase. Controls included incubation of the sections with antisera after fluid phase adsorption with the antigen.

and cytochrome-*c*-oxidase reactive sites were again complementary in layers V and VI. In layer V the moderately cytochrome-*c*-oxidase-rich puffs (Fig. 3c) were calbindin-negative (Fig. 3d). Cytochrome-*c* oxidase reactivity was high in layer upper VI and moderate in lower VI, where puffs have been detected (Figs 2b, 3b, c). Only some calbindin-positive neurones occurred in layer VI and the white matter (Fig. 3a).

Calbindin-positive perikarya could be divided into two classes according to their staining intensity. The larger population consisted of moderately immunoreactive neurones and was found in layers II, III, IVa and the upper halves of layers IVc- α and IVc- β (Figs 2a, 3a, e). The immunostaining of these non-pyramidal cells was mostly limited to the perikaryon and the proximal part of the dendrites. A second smaller population consisted of strongly immunoreactive non-pyramidal cells which were totally labelled and were preferentially situated in layers II, V and VI.

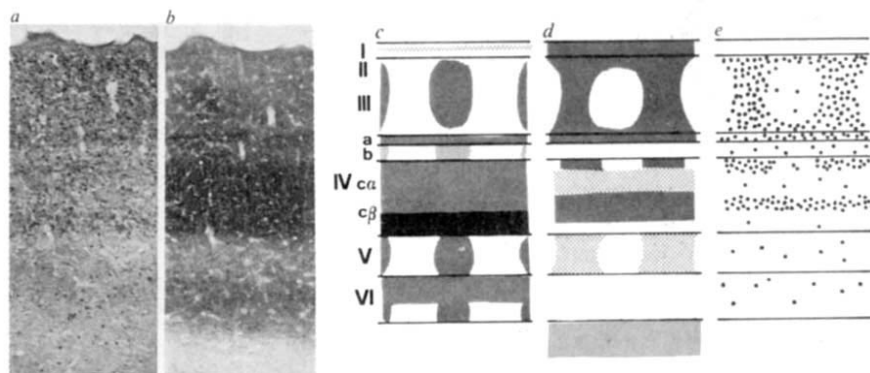
Calbindin-positive axons and terminals were not easily resolved in the neuropil of area 17. Nevertheless, both profuse and diffuse labelling was evident in layers I and V, in addition to those layers in which the moderately immunoreactive cell bodies were located (layers II–III, IVa, upper IVc- α , upper IVc- β ; Fig. 3d). A few coarse and straight calbindin-positive axons were seen tangentially oriented at the boundary between the grey and white matter and some ran obliquely in the grey matter. The white matter contained many extremely thin calbin-

din-positive axons—barely resolved by light microscopy—which occupied the stratum sagittale externum (Fig. 2a, 3d).

The most remarkable characteristic of the immunostaining with calbindin antisera in the squirrel monkey visual cortex was the pervasive, although moderate, labelling of the neuropil and of a large number of neurones in some layers. A new parcellation in layers III, IVc- α and V and a further sublamination of layer IVc emerged (Fig. 3d, e).

Area 17 contained a pattern of strong staining with calbindin antisera which did not coincide with parts that receive dense geniculate terminals. This correlates with the observation that in the simian dorsal lateral geniculate nucleus there are few, if any, calbindin-immunoreactive projection neurones. Rather the lamination seen with calbindin antisera in the visual cortex corresponds to the superposition of nonspecific thalamic^{10,11} and subcortical^{11,12} afferent systems, cortico-cortical association systems^{11,13} and intracortical connections¹¹. In fact, most neuronal perikarya of the systems giving rise to these 'non-geniculate' afferents to area 17 contain strong calbindin immunoreactivity^{14,15}. On the other hand, the cytochrome-*c*-oxidase-rich geniculocortical compartment contains parval-

Fig. 3 *a, b*, Detection of calbindin (*a*) and cytochrome oxidase (*b*) in two consecutive sections of the squirrel monkey visual cortex (magnification of a zone close to the four arrow-heads in Fig. 2). Layer IVc is divided into four sublaminae by the calbindin immunolabelling. Layer V is cytochrome oxidase poor but shows a faint calbindin-positive neuropil whereas the reverse is true for layer VI. $\times 30$. *c-e*, Schematic drawing of the staining pattern observed with the calbindin antisera compared with cytochrome oxidase histochemistry (*c*). Calbindin-positive neuropil (*d*) and perikarya (*e*) occur in the same layers with the exception of layers I, IVb and VI. Note the complementarity of the calbindin-positive neuropil and cytochrome-rich sites in layers I, II-III, V and VI. The drawing is to scale and represents observations on immunohistochemically treated sections cut in various planes of the visual cortex of six animals. No distinction is made between strongly and moderately immunostained interneurons. All calbindin-positive neurones in the primary visual cortex had round cell bodies and diameters of $\sim 10 \mu\text{m}$ in layer II and $20 \mu\text{m}$ in layers V and VI. The strongly stained interneurons of layer II formed bundles of cell processes which descended perpendicular to the pia mater to lower layers. The cytochrome oxidase diagram is based on our own data and on the literature (for example, refs 9, 20). $\times 30$.



bumin (M.R.C., unpublished observation), another calcium-binding protein of the buffer/transport type¹⁶. Thus, geniculocortical and extra-geniculocortical compartments differ in their content of soluble calcium-binding proteins.

Calbindin-rich, cytochrome-*c*-oxidase-poor regions in the 'non-puff' areas may harbour cells that have lower spontaneous activity and well-developed orientation selectivity¹⁷. However,

Ca^{2+} involvement in the expression of these characteristics has not been studied, and the significance of calbindin in the function of such neurones remains unknown. Although intracellular calbindin could act as an additional modulator of the activity of these neurones⁷ it might also be involved in the control of their metabolism and cytoskeletal integrity.

Our studies suggest that the calcium-binding protein calbindin represents a marker for a system which identifies non-geniculate connections (and intracortical connections) of the primary visual cortex. Particularly in the supra- and infragranular layers of the primary visual cortex, the non-geniculate and geniculate systems are segregated into different compartments. This parallel organization is probably integrated only in hierarchically higher visual areas of the cortex.

We thank C. Hemmerle, D. Oeschger and H. Weber for technical and photographic assistance, Dr H. Weber (Sandoz) for providing the monkey for Fig. 1, and Drs J. Lund (Pittsburgh), C. Blakemore (Oxford) and A. Hendrickson (Seattle) for their interest and for comments on the manuscript. The work was supported by grants NF 3.559.0.83, Hartmann Müller-, Mobiliar-, Rentenanstalt-, Sandoz- and EMDO-Stiftung and USPHS grant AM-09012-21. This study was started in 1982 while M.R.C. was a postdoctoral fellow (supported by the Swiss National Foundation) in the laboratory of F.E.B., then at the A. V. Davis Center for Behavioral Neurobiology, Salk Institute, San Diego, California 92138, USA. An abstract of our results has been published elsewhere¹⁸.

Received 20 February; accepted 28 July 1986.

1. Kretschmer, R. H. *Neurosci. Res. Prog. Bull.* **19**, 213-328 (1981).
2. Singer, W. *Vision Res.* **25**, 389-396 (1985).
3. Wassermann, R. H. in *Vitamin D* (eds Norman, A. W., Schaefer, K., Grigolet, H. G. & Herrath, D.) 321 (De Gruyter, Berlin, 1985).
4. Wassermann, R. H. & Taylor, A. N. *Science* **152**, 791 (1966).
5. Norman, A. W., Roth, J. & Orci, L. *Endocr. Rev.* **3**, 331-366 (1982).
6. Wassermann, R. H. & Fullmer, C. S. in *Calcium and Cell Function* (ed. Cheung, W.) 175 (Academic, New York, 1982).
7. Baimbridge, K. G. & Miller, J. J. *Brain Res.* **324**, 85-90 (1984).
8. Christakos, S., Friedlander, E. J., Frandsen, B. R. & Norman, A. W. *Endocrinology* **104**, 1495 (1979).
9. Carroll, E. W. & Wong-Riley, M. T. T. *J. comp. Neurol.* **222**, 1 (1984).
10. Macchi, G. & Bentivoglio, M. *Ital. J. Neurol. Sci.* **2**, 83-96 (1982).
11. Tigges, J. & Tigges, M. in *Cerebral Cortex Vol. 3* (eds Peters, A. & Jones, E. G.) 351-378 (Plenum, New York, 1985).
12. Doty, R. W. *J. comp. Neurol.* **218**, 159 (1983).
13. Van Essen, D. C. in *Cerebral Cortex Vol. 3* (eds Peters, A. & Jones, E. G.) 259 (Plenum, New York, 1985).
14. Celio, M. R. & Norman, A. W. *Anat. Embryol.* **173**, 143-148 (1985).
15. Garcia-Segura, L. M., Baetens, D., Roth, J., Norman, A. W. & Orci, L. *Brain Res.* **296**, 75-86 (1984).
16. Heizmann, C. W. *Experientia* **40**, 910-921 (1984).
17. Hubel, D. *Nature* **299**, 515-524 (1982).
18. Celio, M. R., Schäfer, L., Morrison, J. H., Norman, A. W. & Bloom, F. E. *Soc. Neurosci. Abstr.* **10**, 1 (1983).
19. Wong-Riley, M. *Brain Res.* **171**, 11 (1979).
20. Livingstone, M. S. & Hubel, D. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6098 (1982).

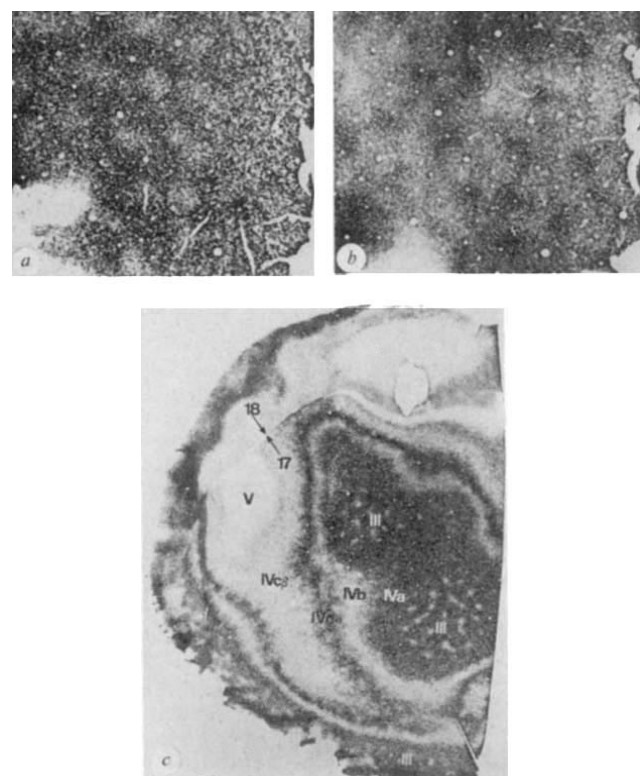


Fig. 4 *a*, Tangential section mainly through layer III of the convex surface of areas 17 of a squirrel monkey. The calbindin immunolabelling is discontinuous and shows lacunae of diminished immunoreactivity (some marked by circles). Animal 29/85. $\times 40$. *b*, Consecutive section to *a* reacted for the detection of cytochrome-*c*-oxidase-reactive sites. Cytochrome-oxidase-rich puffs (some marked by circles) correspond exactly to the calbindin lacunae in layer III of Fig. 3a. $\times 40$. *c*, Tangential section grazing layer III and comprising layers IVa, IVb, IVc- α and IVc- β . Note the calbindin lacunae in layer III and similar inhomogeneities in the staining of upper layer IVc- α . The lacunae were more obvious in this animal than in the monkey shown in *a*. Animal 81/85. $\times 30$.

Single apamin-blocked Ca-activated K^+ channels of small conductance in cultured rat skeletal muscle

Andrew L. Blatz & Karl L. Magleby

Department of Physiology and Biophysics, University of Miami
School of Medicine, Miami, Florida 33101, USA

Action potentials in many excitable cells are followed by a prolonged afterhyperpolarization¹⁻³ that modulates repetitive firing^{2,4,5}. Although it is established that the afterhyperpolarization is produced by Ca-activated K^+ currents^{1,2,4}, the basis of these currents is not known. The large conductance (250 pS) Ca-activated K^+ channel (BK channel^{6,7}) is not a major contributor to the afterhyperpolarization in non-innervated skeletal muscle and some nerve cells, because apamin, a neurotoxic component of bee venom, abolishes the afterhyperpolarization⁸⁻¹¹ but does not block BK channels⁹, and 5 mM extracellular tetraethylammonium ion (TEA) blocks BK channels^{9,12} but does not reduce the afterhyperpolarization^{2,4,8-11}. We now report single-channel currents from small conductance (10–14 pS) Ca-activated K^+ channels (SK channels) with the necessary properties to account for the afterhyperpolarization. SK channels are blocked by apamin but not by 5 mM external TEA (TEA_o). They are also highly Ca-sensitive at the negative membrane potentials associated with the afterhyperpolarization.

Figure 1a shows activation of SK channels by Ca^{2+} at the inner membrane surface (Ca_i) of an excised patch of muscle membrane. In the absence of Ca_i ($<10^{-9}$ M free Ca^{2+}) there was no channel activity. Increasing Ca_i to 0.1 μ M at the downward arrow activated several SK channels after a delay resulting from the solution-exchange time. Channel activity ceased after Ca_i was returned to zero at the upward arrow in Fig. 1b. Ca^{2+} at the outer membrane surface did not activate SK channels.

The amplitudes of single-channel currents through SK channels increased linearly with membrane potential over the examined range of -20 to -60 mV (Fig. 1c–e), giving a single-channel slope conductance of ~ 12 pS with symmetrical 140 mM KCl. Reversing the driving force on K^+ with positive membrane potentials resulted in outward currents through SK channels.

Compared with BK channels, which are highly voltage-dependent¹³⁻¹⁵, the activity of SK channels had little voltage sensitivity. SK channels were typically active at all membrane potentials examined (-90 to $+40$ mV) with 0.1–1 μ M Ca_i , and the net current through SK channels in a membrane patch activated by 0.1 μ M Ca_i at $+40$ mV was 80% of the current observed at -40 mV for the same Ca_i .

Inward currents through SK channels were no longer observed when K^+ in the external solution was replaced by Na^+ , and conversely, outward currents were no longer observed when K^+ in the internal solution was replaced by Na^+ . Replacing internal or external Cl^- with methyl sulphate had no effect on current through SK channels. These results indicate that SK channels are selective for K^+ over Na^+ and Cl^- .

In addition to the SK and BK channels, a very small conductance (4 pS) Ca-activated channel was also observed occasionally, with 0–2 channels in a patch. The closing of this channel following a return to 0 Ca_i is indicated by the asterisk in Fig. 1b.

A typical membrane patch contained 10–60 SK channels, but some patches had none. Most membrane patches also contained 1–5 BK channels, allowing the determination of the relative Ca-sensitivity of these two types of channels under identical conditions. When Ca_i was increased progressively from zero to 1 μ M (Fig. 2), SK channels were activated before BK channels. (Returning Ca_i to zero reversed the sequence.) This sequence was always observed at negative membrane potentials. These

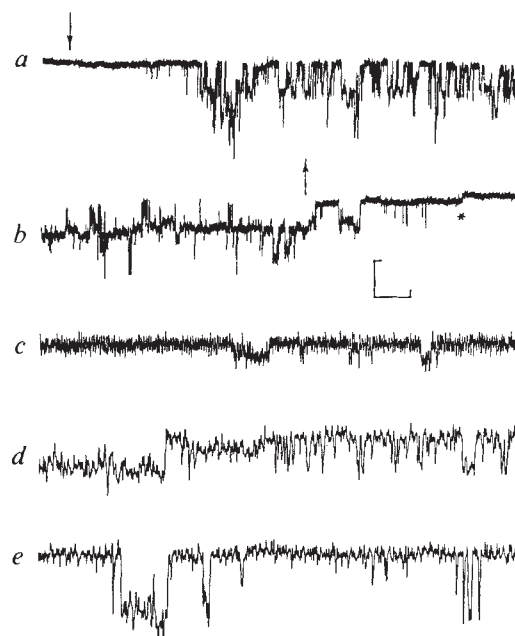


Fig. 1 Single-channel currents through SK channels. Downward step changes in current indicate channel opening (inward currents). *a*, Activation of SK channels by Ca_i . At the downward arrow Ca^{2+} in the solution perfusing the inner membrane surface was increased from 0 to 0.1 μ M free Ca . Up to three of the 10 SK channels in the membrane patch could be active simultaneously in 0.1 μ M free Ca . Membrane potential, -40 mV. *b*, Reversibility of Ca-activation of SK channels. Continuation of the experiment in *a* as Ca_i was reduced at the upward arrow from 0.1 μ M to 0. Asterisk, closing of very small conductance (4 pS) Ca-activated channel. *c–e*, Currents through SK channels at -20 , -40 and -60 mV, respectively. Horizontal bar, 1.5 s in *a* and *b*, 250 ms in *c* and 100 ms in *d* and *e*; vertical bar, 0.5 pA in *a–e*. Because of the filtering, the apparent single openings in *a–e* probably represent bursts of openings with a mean burst duration of ~ 25 ms at -40 mV.

Methods. Single-channel currents were recorded from excised inside-out membrane patches from surface membranes of tissue cultured rat skeletal muscle using standard techniques²⁴. Solutions were changed using a micropuffing chamber¹³. Unless noted, all solutions contained (mM): 140 KCl, 5 TES buffer, 0.5 EGTA and sufficient added $CaCl_2$ to provide the stated free $[Ca^{2+}]$. Ca/EGTA ratios for desired free Ca as in ref. 13. Solutions referred to as zero Ca contained $<10^{-9}$ M free Ca^{2+} . Membrane potentials defined conventionally, as the potential at the inside membrane surface minus the potential at the outside membrane surface. Experiments performed at 20–22 °C. Single-channel currents recorded on FM tape and played at reduced speed into a chart recorder with effective low-pass filtering of 120 Hz (-3 dB, Bessel, 24 dB per octave).

observations, together with the absence of BK channel activity in the presence of SK channel activity with 10-fold less Ca_i at -40 mV (Fig. 1), indicates that SK channels are more Ca-sensitive than BK channels at negative membrane potentials. At $+40$ mV the Ca-sensitivity of the two channels was similar, as the Ca-sensitivity of BK channels increases with depolarization¹³⁻¹⁵.

An amplitude histogram obtained from the current during inflow of Ca_i in Fig. 2a,b, but before the BK channels became active, is plotted in Fig. 2d. The measured current fluctuated around five discrete levels which correspond to currents for 0, 1, 2, 3 or 4 channels open at one time, and are separated by 0.5 pA, the same amplitude as the step changes in current in Fig. 1a for the opening of an SK channel. Similar 0.5-pA spacing was observed during washing out of Ca_i . These results indicate that the current during the addition and removal of Ca_i when the BK channels are not active is composed of summed unitary currents of identical amplitude to those of the SK channel.

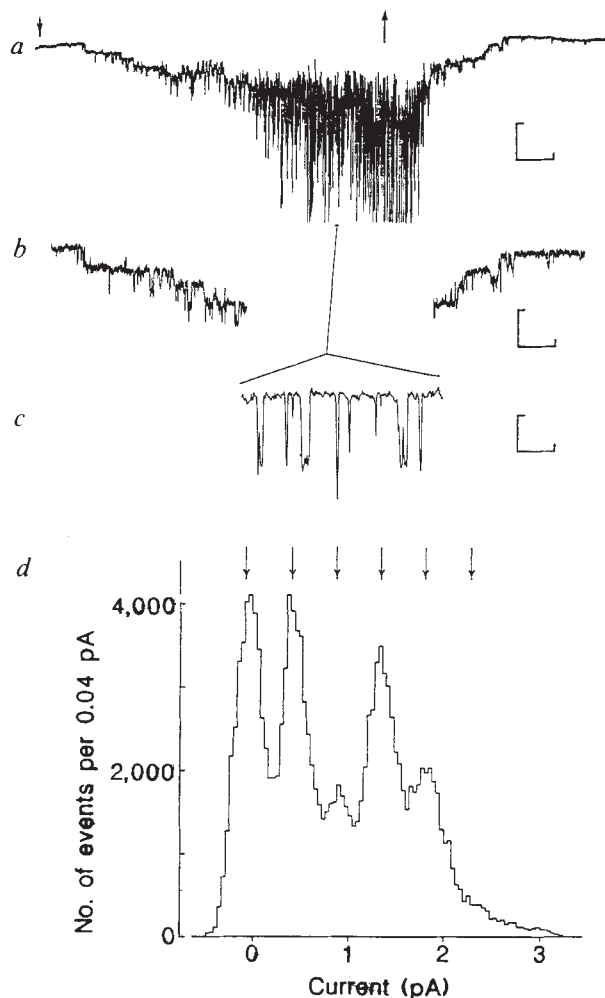


Fig. 2 Ca-activation of SK and BK channels in the same membrane patch. *a*, Ca_i raised from 0 to $1 \mu\text{M}$ at the downward arrow and then returned to zero at the upward arrow. Small and large downward current steps arise from SK and BK channels, respectively. Currents through BK channels are truncated because of filtering and limits of chart recorder. *b*, Currents through SK channels recorded at higher gain as Ca_i was raised (left) and lowered (right). *c*, BK channel currents at lower gain and less filtering (1.5 kHz, -3 dB) superimposed on summed current (noisy baseline) from ~ 5 -7 SK channels. Membrane potential, -40 mV . The membrane patch contained 6 BK channels open $\sim 2\%$ of the time in $1 \mu\text{M}$ Ca_i . Vertical bar (pA): *a*, 2; *b*, 1; *c*, 5. Horizontal bar (s): *a*, 5; *b*, 2.5; *c*, 0.05. *d*, Amplitude histogram of the currents during onset of the SK-channel activity shown in *a* and *b*. Arrows 0.5 pA apart, the current through a single SK channel. The third peak appears low because the data were collected under non-stationary conditions while Ca_i was increasing. Consequently, the peak amplitudes would not be expected to follow a binomial distribution, even if the channels functioned independently of one another. Amplitude histograms were obtained by measuring the current each millisecond by computer, sorting the current amplitudes into bins according to amplitude, and plotting the number of events in each bin against the mean current for the bin.

The Ca -sensitivity of the SK channel was determined by measuring Ca -activated currents through membrane patches containing 30-60 SK channels. BK channels were blocked with 5 mM TEA in the patch pipette^{9,12}. TEA_o had little or no effect on the SK channels. Stepping Ca_i from 0 to $1 \mu\text{M}$ led to a large increase in current and fluctuation of current about the mean, consistent with 47 channels open at the peak current (Fig. 3*a*). Steps in Ca_i from 0 to 0.1 or 0.05 μM produced less mean current and less fluctuation of the current about the mean (Fig. 3*b, c*). Higher gain (not shown) indicated that with 0.05 μM Ca_i , only

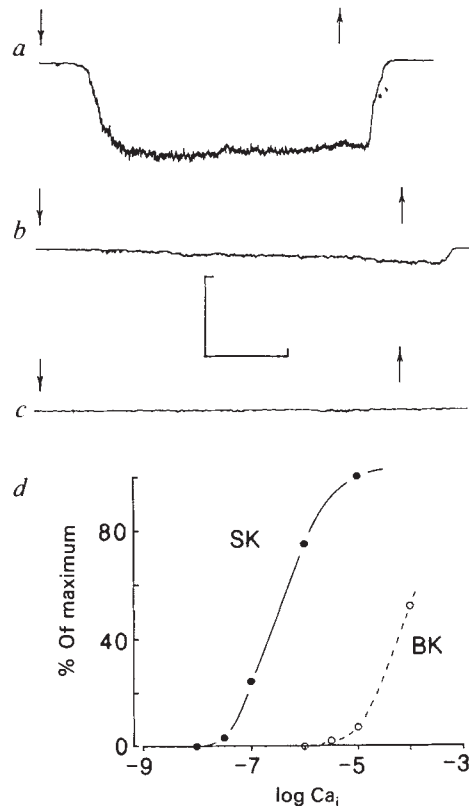


Fig. 3 Ca -sensitivity of SK channels. *a-c*, Ca -activated currents through SK channels. Ca_i was stepped from zero to $1 \mu\text{M}$ (*a*), $0.1 \mu\text{M}$ (*b*) or $0.05 \mu\text{M}$ (*c*) at the downward arrows and then back to zero at the upward arrows. The external solution also contained 5 mM TEA. Potential, -40 mV . *a-c*; Vertical bar, 20 pA; horizontal bar, 20 s. *d*, Per cent maximum response as a function of Ca_i for SK channels (filled circles) and per cent open time as a function of Ca_i as determined previously¹³ for the BK channel in myotubes at -40 mV (open circles).

three of the channels in the patch were active at any one time. Figure 3*d* plots the results of the activation of SK channels by Ca_i at -40 mV (filled circles). Data obtained previously¹³ for the BK channel at the same membrane potential are also presented (open circles). The much greater Ca -sensitivity of SK channels (half-maximal response 2 - $5 \times 10^{-7} \text{ M}$ Ca_i) than BK channels in cultured muscle is apparent, although the Ca -sensitivity of BK channels in cultured muscle is less than in some other tissues¹⁶.

Fifteen minutes after injecting apamin into the patch pipette, where it could diffuse to the extracellular membrane surface, currents through SK channels were largely abolished whereas BK channels were still active (Fig. 4*a, b*). Examination of records at faster time-base confirmed no obvious effect of apamin on BK channels⁹. After BK channels were blocked with 5 mM TEA in the patch pipette^{9,12} the apamin block of current through SK channels was seen more clearly (Fig. 4*c, d*). SK channels could also be blocked by filling the recording pipette with solutions containing $<1 \mu\text{M}$ apamin before sealing to the membrane.

When BK and SK channels were blocked with TEA and apamin, a small Ca -activated current was still present (Fig. 4*d*). Inspection of this Ca -activated current at higher gain and for several different Ca_i suggests poorly resolved 0.1 - 0.2 pA steps (Fig. 4*e*) similar to those of the very small conductance Ca -activated channel. We cannot exclude, however, that the apamin and TEA-resistant Ca -activated current arises from possible leakage of K^+ through apamin-blocked SK channels. Leakage of K^+ through TEA-blocked BK channels is unlikely as BK channels would not have opened at the Ca_i and membrane

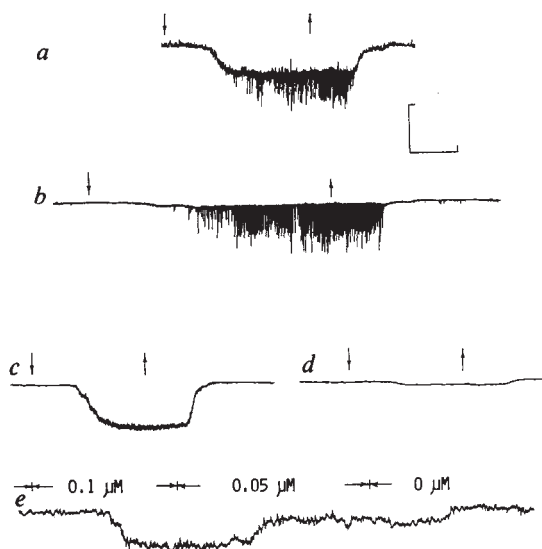


Fig. 4 Apamin block of SK channels. *a-d*, Ca_i was stepped from zero to $1 \mu\text{M}$ at the downward arrows and back to zero at the upward arrows. Currents through a membrane patch containing SK and BK channels before (*a*) and 15 min after (*b*) injection of a small amount of solution containing $100 \mu\text{M}$ apamin into the patch pipette. *c, d*, Apamin block of SK channels after TEA block of BK channels with 5 mM TEA in the patch pipette. *c*, Before apamin; *d*, after apamin. Ca-activated 0.5-pA current steps consistent with SK channels (as in Fig. 2) were observed before application of apamin and never after. *e*, Residual apamin- and TEA-resistant Ca-activated current (same experiment as in *c, d*) shown at higher gain as Ca_i was raised from 0 to $0.1 \mu\text{M}$, lowered to $0.05 \mu\text{M}$ and then returned to 0 . Potential, -40 mV . Vertical bar: *a-d*, 10 pA ; *e*, 1 pA ; horizontal bar: *a-d*, 20 s ; *e*, 10 s . Apamin was added to the solution in the patch pipette by inserting a thin plastic tube pulled from a syringe deep into the patch pipette (within $<1 \text{ mm}$ from the attached membrane patch) and injecting the toxin.

potential of the experiment in Fig. 4*e*. Apamin does not block Ca-activated K^+ fluxes in erythrocytes¹⁷, which may originate from a channel with conductance of 18 pS (ref. 18). It is not known whether apamin blocks a Ca-activated 19-pS channel in snail neurones¹⁹.

Our findings indicate that SK channels satisfy the criteria expected for channels underlying the prolonged afterhyperpolarization in cultured skeletal muscle: they are blocked by apamin and not by 5 mM TEA, and they are highly Ca-sensitive at negative membrane potentials. SK channels may represent at least some of the apamin receptors found in brain and various other mammalian tissues which are modulated by an endogenous apamin-like substance²⁰, as the physiological effects of apamin block are consistent with a reduction of Ca-activated K^+ permeability^{8,10,21,22}. SK channels may also represent the apamin receptor found in muscle membrane of patients with myotonic muscular dystrophy²³.

This work was supported by grants from NIH (AM32085) and the Muscular Dystrophy Association.

Received 14 April; accepted 25 July 1986.

- Meech, R. A. *Rev. Biophys. Bioengng* **7**, 1-18 (1978).
- Barrett, J. N., Barrett, E. F. & Dribin, L. B. *Dev. Biol.* **82**, 258-266 (1981).
- Schmid-Antomarchi, H. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 2188-2191 (1985).
- Barrett, J. N. & Barrett, E. F. *J. Physiol., Lond.* **255**, 737-774 (1976).
- Kawai, T. & Watanabe, M. *Br. J. Pharmac.* **87**, 225-232 (1986).
- Marty, A. *Nature* **291**, 497-500 (1981).
- Pallotta, B. S., Magleby, K. L. & Barrett, J. N. *Nature* **293**, 471-474 (1981).
- Hugues, M., Romey, G., Duval, D., Vincent, J. P. & Lazdunski, M. *Proc. natn. Acad. Sci. U.S.A.* **79**, 1308-1312 (1982).
- Romey, G. & Lazdunski, M. *Biochem. biophys. Res. Commun.* **118**, 669-674 (1984).
- Romey, G., Hughes, M., Schmid-Antomarchi, H. & Lazdunski, M. *J. Physiol., Paris* **79**, 259-264 (1984).
- Pennefather, P., Lancaster, B., Adams, P. R. & Nicoll, R. A. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3040-3044 (1985).
- Blatz, A. L. & Magleby, K. L. *J. gen. Physiol.* **84**, 1-23 (1984).
- Barrett, J. N., Magleby, K. L. & Pallotta, B. S. *J. Physiol., Lond.* **331**, 211-230 (1982).

- Moczydlowski, E. & Latorre, R. *J. gen. Physiol.* **82**, 511-542 (1983).
- Methfessel, C. & Boheim, G. *Biophys. Struct. Mech.* **9**, 35-60 (1982).
- Petersen, O. H. & Maruyama, Y. *Nature* **307**, 693-696 (1984).
- Burgess, G. M., Claret, M. & Jenkinson, D. H. *J. Physiol., Lond.* **317**, 67-90 (1981).
- Hamill, O. P. *J. Physiol., Lond.* **319**, 97-98P (1981).
- Lux, H. D., Neher, E. & Marty, A. *Pflügers Arch. ges. Physiol.* **389**, 293-295 (1981).
- Fosset, M., Schmid-Antomarchi, H., Hugues, M., Romey, G. & Lazdunski, M. *Proc. natn. Acad. Sci. U.S.A.* **81**, 7228-7232 (1984).
- Banks, B. E. C. *et al. Nature* **282**, 415-417 (1979).
- Habermann, E. *Pharmac. Theor.* **25**, 255-270 (1984).
- Renaud, J. *et al. Nature* **319**, 678-680 (1986).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. *Pflügers Arch. ges. Physiol.* **391**, 85-100 (1981).

Strategy by which nitrogen-fixing unicellular cyanobacteria grow photoautotrophically

A. Mitsui, S. Kumazawa, A. Takahashi, H. Ikemoto, S. Cao & T. Arai

Division of Biology and Living Resources, School of Marine and Atmospheric Science, University of Miami, Miami, Florida 33149, USA

Among nitrogen-fixing microorganisms, nitrogen-fixing cyanobacteria are unique in their ability to carry out oxygen-evolving photosynthesis and oxygen-labile nitrogen fixation within the same organisms¹⁻³. These seemingly incompatible reactions take place in heterocystous cyanobacteria by the spatial separation of the site of nitrogen fixation (heterocysts) from the site of photosynthesis (vegetative cells)^{4,5}. Several hypotheses have been proposed to explain these mechanisms in non-heterocystous cyanobacteria^{3,6-11}. Using batch cultures of *Gloeotheca (Gloeocapsa) spp.*, Gallon and collaborators demonstrated the mechanism of temporal separation of photosynthesis and nitrogen fixation into the light and dark periods of growth, respectively⁹. However, the mechanisms by which these two incompatible reactions can occur under continuous light conditions still remained ambiguous. Using novel strains of aerobic nitrogen-fixing, unicellular marine cyanobacteria, *Synechococcus spp.*, grown under synchronized conditions, we report here that nitrogen fixation and photosynthesis occur at different phases in the cell division cycle. Our data, obtained under both diurnal light/dark cycle and continuous illumination, indicate that the temporal separation of the two phases during the cell division cycle is the mechanism by which these unicells can grow photoautotrophically under nitrogen-fixing conditions.

To investigate the strategy by which unicellular cyanobacteria grow photoautotrophically under nitrogen-fixing conditions, we examined whether exponentially growing cells separate the phases of photosynthesis and nitrogen fixation within their cell cycle. Under batch or continuous culture, such temporally separated events are obscured by the activities of randomly phased cells; hence the two apparently incompatible reactions appear to occur concurrently. Provided that the temporal separation is the key factor, the cell-cycle events can be more clearly seen under synchronized conditions. Until recently, however, suitable strains have not been available for this kind of experimentation with synchronous cultures. We have isolated new unicellular aerobic nitrogen-fixing strains (*Synechococcus spp.* Miami BG 43511 and 43522) (Fig. 1) suitable for this work from our extensive survey of marine photosynthetic microorganisms¹²⁻¹⁴.

Synchronized growth of strains 43511 and 43522 under aerobic nitrogen-fixing conditions was initiated by the interruption of exponential growth by darkness. Figure 1 shows the cells during and after the cell division under synchronized growth. Synchronized growth was maintained under both conditions with 3 cycles of 12 h light/12 h dark (12L/12D) and subsequently with 20 h continuous illumination (Fig. 2). Although the increase in cell numbers is not always twofold, cell division occurs syn-

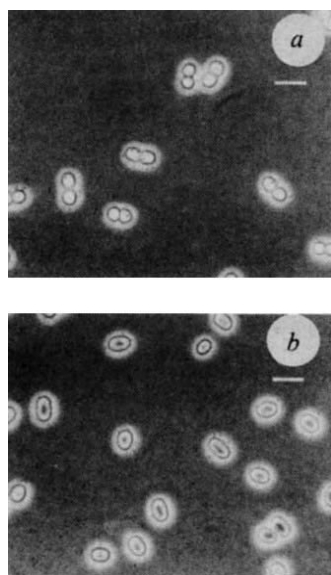


Fig. 1 Phase-contrast photomicrographs of the aerobic nitrogen-fixing marine unicellular cyanobacterium *Synechococcus* sp. strain Miami BG 43511. Cells are shown during the division period (a) and the elongation period (b) under aerobic diazotrophic synchronous growth conditions. Bar, 10 μm . Cell size of *Synechococcus* sp. strain Miami BG 43522 is slightly smaller than strain 43511. One cell division cycle (a-b-a) took ~ 20 h at 30°C, pH 7.6 in a photon radiance of $150 \mu\text{mol m}^{-2} \text{s}^{-1}$ and aeration with 4% CO_2 in air. Such cell division cycles (a-b-a) were repeated during synchronous growth as shown in Fig. 2c.

chronously as indicated by the stepwise increase in cell numbers (Fig. 2c, solid line). The synchronicity of the culture is further indicated by the transient appearance of doubling cells (Fig. 1a) only during a well-segregated period of the cell cycle (Fig. 2c, dotted line). During stepwise growth, net oxygen evolution (Fig. 2a, solid line) and nitrogenase (Fig. 2a, dotted line) activities appear and disappear in a reciprocal manner. During

the cycles of 12L/12D conditions, maximum net oxygen evolution activity appears just before the cell division. The oxygen evolution activity decreased to almost zero (first and second L/D cycles) or below zero (third L/D cycle) and increased again to form one cycle of change within a cell cycle. Appearance of nitrogenase activity was observed when the net oxygen evolution activity decreased to a certain level. The maximum activity of nitrogenase always coincided with the minimum activity of net oxygen evolution capability.

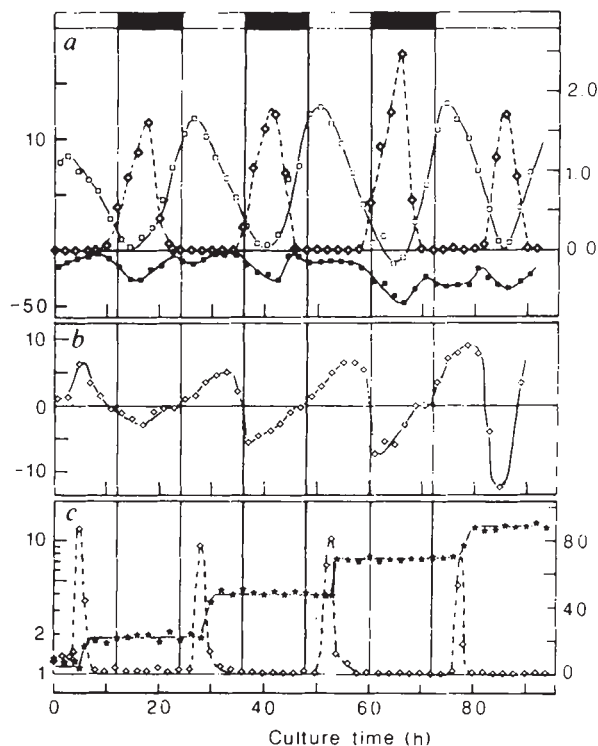
The oscillation of photosynthetic net oxygen evolution activity is observed even under continuous illumination started after 72 h (Fig. 2a). Again, the appearance of nitrogenase activity coincides with the phase of declining photosynthetic oxygen evolution capability and the maximum nitrogenase activity is observed when the oxygen evolution capability is minimum (Fig. 2a—note especially the section after 72 h).

The activity of oxygen uptake in the dark also changes during cell growth (Fig. 2a, dark respiration). In the case of dark oxygen uptake activity, the periodicity of the change is ambiguous although the dark oxygen uptake activity always seems higher when nitrogenase activity is higher. Respiratory activity has been suggested as one mechanism by which the anaerobicity for nitrogen fixation is maintained¹⁵. However, as shown in Fig. 2a, the magnitude of the changes in oxygen uptake activity are much smaller than those in oxygen evolution activity. Moreover, respiration in the strains from the same genus is inhibited by the light¹⁶. Therefore, large oscillations of photosynthetic oxygen evolution capability cannot be explained solely by the dark respiratory activity, even if it is concurrent with the light.

Note that the oxygen evolution activity shown in Fig. 2a was not obtained *in situ* but in an oxygen-measuring reaction vessel (see Fig. 2 legend). Thus, the change in the oxygen evolution activity is a quantitative index of the photosynthetic potential capability of the cell. In this context, the net changes in carbohydrate content in the culture provide insight into the *in situ* changes in photosynthetic potential. Interestingly, the carbohydrate content in the culture goes through phases of net accumulation and net consumption under both conditions with 12L/12D and continuous illumination (Fig. 2b). It is obvious that carbohydrate consumption proceeds during the dark phase of

Fig. 2 Changes in photosynthetic oxygen evolution, respiratory oxygen uptake and acetylene reduction capabilities (a); in the rate of carbohydrate synthesis (b); and in total cell number and occurrence of doubling cells (c) during the synchronous growth of *Synechococcus* sp. strain Miami BG 43511. a, Left-hand scale, oxygen exchange ($\mu\text{mol O}_2$ per mg dry weight per h); right-hand scale, nitrogenase activity ($\mu\text{mol C}_2\text{H}_2$ reduced per mg dry weight per h). $\square$, O_2 photoproduction; $\circ$, nitrogenase; $\blacksquare$, dark respiration. b, μg glucose equivalent per ml culture per h. c, Left-hand scale, total cell number ($\ast$); right-hand scale, doubling cell ($\diamond$).

Methods. The strain BG 43511 was grown under the same nitrogen-fixing conditions (30°C, photon radiance $150 \mu\text{mol m}^{-2} \text{s}^{-1}$, aeration with 4% CO_2 in air at 200 ml min^{-1}) as described previously²⁴. The culture medium²⁴ did not contain combined nitrogen growth nutrients. For the induction of synchronous growth, an early exponential phase culture ($\sim 10^6$ cells per ml) was interrupted during growth by a 20-h dark period without aeration. After the dark induction period, synchronized growth was monitored during the light and dark periods as indicated. During the dark, the aeration was stopped but the culture was mixed by magnetic stirrer. Continuous illumination was started after 72 h. Aliquots were taken every 2 h and assayed for oxygen exchange and acetylene reduction capabilities, dry weight, carbohydrate content and cell numbers. Oxygen exchange was monitored using a Clark-type electrode in a reaction vessel as described previously²⁵. The sample (2 ml) was transferred from the culture cylinder into the reaction vessel without any treatment and bubbled with 4% CO_2 in air in the dark for 2–3 min for equilibration (30°C). After the equilibration in the reaction vessel, the rates of respiration and photosynthetic oxygen evolution were measured in the dark and the light (photon radiance $2,000 \mu\text{mol m}^{-2} \text{s}^{-1}$), respectively. The light intensity for oxygen evolution was within a light saturation range for this strain (photon radiance $1,000$ – $4,000 \mu\text{mol m}^{-2} \text{s}^{-1}$) under the given experimental conditions. The acetylene reduction (10% C_2H_2 in air, photon radiance $150 \mu\text{mol m}^{-2} \text{s}^{-1}$ at 30°C), cell number and dry weight were measured as described previously²⁴. Carbohydrate content was measured by the phenol-sulphuric acid method²⁶. Rates of carbohydrate synthesis were calculated from the changes in carbohydrate content (per ml culture basis) measured every 2 h.



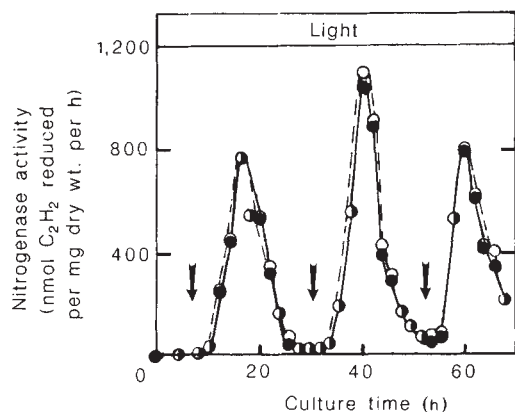


Fig. 3 Effect of DCMU on the nitrogenase activity of synchronously growing *Synechococcus* sp. Miami BG 43522 under continuous illumination. Synchronous growth was induced by the insertion of dark-light-dark treatment of 20 h each and then growing the synchronized culture for 70 h under continuous illumination. Samples were taken every 2 h and acetylene-reduction activities were measured under anaerobic conditions (10% C_2H_2 in argon, photon radiance $150 \mu\text{mol m}^{-2} \text{s}^{-1}$ at 30°C) as described previously²⁴ with (open circles) and without (filled circles) DCMU. The final concentration of DCMU was $20 \mu\text{M}$, which eliminated the photosynthetic oxygen evolution of this strain completely (data not shown). Arrows, stage when cell division occurred.

growth¹⁷, but net consumption of carbohydrate also occurs during continuous illumination. Because the consumption of endogenous carbohydrate exceeds its rate of synthesis even under the light condition (Fig. 2b), it can be concluded that during the phase of net carbohydrate consumption there is little net oxygen evolution under *in situ* culture conditions. Moreover, most of the nitrogenase activity appears only when the culture is in the phase of net carbohydrate consumption (Fig. 2a,b), which is consistently observed over at least three cell cycles under continuous illumination¹⁸.

Endogenous carbohydrate (glycogen) has been considered as an ultimate source of reductant for nitrogenase catalysis in some cyanobacteria^{13,19-22}. We therefore suspect that endogenous carbohydrate is important in the nitrogen fixation in strains 43511 and 43522. The oscillating oxygen evolution and carbohydrate synthesis, together with the timely appearance of nitrogenase activity, strongly indicate that there is a temporal separation of photosynthesis and nitrogen fixation within a cell cycle of the unicellular aerobic nitrogen-fixing *Synechococcus* strains.

In the non-heterocystous filamentous cyanobacterium *Oscillatoria* sp., there is a ~threefold stimulation of acetylene reduction activity when photosynthetic oxygen evolution is completely inhibited by dichlorophenyl dimethylurea (DCMU), an inhibitor of photosynthetic oxygen evolution²³. This observation indicates that photosynthetic oxygen evolution has adverse effects on nitrogen fixation. In the case of *Synechococcus* spp. described here, there may be little or no photosynthetic oxygen evolution during the phase of nitrogen fixation provided that temporal separation is functional. If our interpretation is correct, DCMU should have no effect on the acetylene reduction activity. As shown in Fig. 3, the activity of acetylene reduction during three-cell division cycles under continuous light is not affected by the addition of DCMU at concentrations that abolish photosynthetic oxygen evolution. This can be taken as additional evidence to support the idea of temporal separation.

Supported by NSF grants to A. Mitsui (Nos PCM8202429 and CPE 8312092) and the International Ocean Development Foundation. We thank Drs R. H. Burris, J. W. Hastings, J. Richard and H. Senger for suggestions. Also supported by the cooperative research programme between the University of Miami and Tokai University. S.K. and A.T. were on leave from Tokai University.

Received 6 May; accepted 21 August 1986.

1. Stanier, R. Y. & Cohen-Bazire, G. A. *Rev. Microbiol.* **31**, 225-274 (1977).
2. Stewart, W. D. P. A. *Rev. Microbiol.* **34**, 497-536 (1980).
3. Gallon, J. R. *Trends biochem. Sci.* **6**, 19-23 (1981).
4. Haselkorn, R. A. *Rev. Plant Physiol.* **29**, 319-344 (1978).
5. Fay, P. in *Recent Advances in Biological Nitrogen Fixation* (ed. Subba Rao, N. S.) 121-165 (Arnold, London, 1980).
6. Fay, P., Kumar, H. D. & Fogg, G. E. *J. gen. Microbiol.* **35**, 351-360 (1964).
7. Gallon, J. R., LaRue, T. A. & Kurz, W. G. W. *Can. J. Microbiol.* **20**, 1633-1637 (1974).
8. Weare, N. M. & Benemann, J. R. *J. Bact.* **119**, 258-265 (1974).
9. Mullineaux, P. M., Gallon, J. R. & Chaplin, A. E. *FEMS Microbiol. Lett.* **10**, 245-247 (1981).
10. Kallias, T. et al. in *Photosynthetic Prokaryotes: Cell Differentiation and Function* (eds Papageorgiou, G. C. & Packer, L.) 281-302 (Elsevier, New York, 1983).
11. Stal, L. J. & Krumbein, W. E. *Archs Microbiol.* **143**, 67-71 (1985).
12. Mitsui, A. in *Proc. 5th Int. Ocean Dev. Conf. I(B1)*, 29-52 (Seino, Tokyo, 1978).
13. Mitsui, A. et al. *Ann. N.Y. Acad. Sci.* **413**, 514-530 (1983).
14. Mitsui, A. et al. in *Biotechnology and Bioprocess Engineering* (ed. Ghose, T. K.) 119-155 (United India, New Delhi, 1985).
15. Gallon, J. R. & Hamadi, A. F. *J. gen. Microbiol.* **130**, 495-503 (1984).
16. Sandmann, G. & Malkin, R. *Archs biochem. Biophys.* **234**, 105-111 (1984).
17. Smith, A. J. in *The Biology of Cyanobacteria* (eds Carr, N. G. & Whitton, B. A.) 47-85 (University of California Press, Berkeley, 1982).
18. Mitsui, A., Takahashi, A., Ikemoto, H., Cao, S. & Arai, T. *5th Int. Symp. Photosynth. Prokaryotes*, Abstr. 63 (Grindelwald, Switzerland, 1985).
19. Mullineaux, P. M., Chaplin, A. E. & Gallon, J. R. *J. gen. Microbiol.* **120**, 227-232 (1980).
20. Kumazawa, S. & Mitsui, A. *Int. J. Hydrogen Energy* **6**, 339-348 (1981).
21. Ernst, A. & Böger, P. *J. gen. Microbiol.* **131**, 3147-3153 (1985).
22. Chaplin, A. E. & Gallon, J. R. *5th Int. Symp. Photosynth. Prokaryotes Abstr.* 253 (Grindelwald, Switzerland, 1985).
23. Stal, L. J. & Krumbein, W. E. *Archs Microbiol.* **143**, 72-76 (1985).
24. León, C., Kumazawa, S. & Mitsui, A. *Curr. Microbiol.* **13**, 149-153 (1986).
25. Kumazawa, S. & Mitsui, A. *Appl. environ. Microbiol.* **50**, 287-291 (1985).
26. Dubois, M., Gilles, K. A., Hamilton, J. K., Rebers, P. A. & Smith, F. *Analyt. Chem.* **28**, 350-356 (1956).

An IgG autoantibody which inactivates C1-inhibitor

J. Jackson, R. B. Sim*, A. Whelan & C. Feighery

Department of Immunology, St James Hospital and Trinity College, Dublin 8, Ireland

*MRC Immunochemistry Unit, Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Antibodies are considered to play a specific pathogenic role in certain disease states such as myasthenia gravis, Graves' disease and autoimmune haemolytic anaemia¹⁻³. Autoantibodies which interfere with the function of enzyme cascade systems have also been described in diseases such as acquired haemophilia (anti-factor VIII antibodies) and glomerulonephritis (C3 nephritic factor)^{4,5}. The identification of these autoantibodies is crucial to an understanding of the aetiology of such diseases and is also of importance in revealing the inter-relationships of the immune system with other biological pathways. This is the first report of an immunoglobulin G (IgG) autoantibody reactive with C1-inhibitor (C1-Inh), a pivotal inhibitor of the inflammatory response which is known to inactivate proteins of the complement, kinin, fibrinolytic and 'contact phase' systems^{6,7}. This autoantibody was isolated from a patient with a novel variant of acquired angioedema and C1-Inh dysfunction. This finding highlights the involvement of the immune system in the pathogenesis of disorders characterized by the presence of dysfunctional inflammatory response proteins.

The patient, an otherwise healthy 62-year-old man, presented with recurrent oedema of recent onset which was clinically characteristic of angioedema. The initial laboratory investigation of his complement system revealed absence of C1-Inh functional activity (performed according to the method of Sim and Reboul⁷) but normal antigenic levels of this protein (22 mg%: range = 18-41 mg%). The serum levels of C1q (<1 mg%: range = 10-25 mg%) and C4 (<4 mg%: range = 10-45 mg%) were reduced. Normal serum levels of C3, C5 and Factor B were recorded. Complement profiles, including C1-Inh function, performed on the patient's healthy offspring were normal. The age of this patient at presentation, the negative family investigations and the low serum C1q, are not consistent with the

Table 1 Effect of IgG on C $\bar{I}$ -Inh function

	ΔA
(i) C $\bar{I}$ s + ZLNE	71
(ii) C $\bar{I}$ s + C $\bar{I}$ -Inh + ZLNE	4
(iii) C $\bar{I}$ s + IgG IEC + C $\bar{I}$ -Inh + ZLNE	48 (6)
C $\bar{I}$ s + IgG Prot. A + C $\bar{I}$ -Inh + ZLNE	41 (4)
C $\bar{I}$ s + IgG IEC:G200 + C $\bar{I}$ -Inh + ZLNE	33 (6)
(iv) C $\bar{I}$ s + IgG SBTI + C $\bar{I}$ -Inh + ZLNE	44
(v) C $\bar{I}$ s + IgG + ZLNE	72
(vi) IgG + ZLNE	2

This assay is based on the ability of purified C $\bar{I}$ s to catalyse a colorimetric conversion of the substrate *N*-carbobenzoyloxy-L-lysine-p-nitrophenyl ester (ZLNE)⁷. Briefly, the protocol used for this assay was as follows: 100 μ l of the purified IgG (9 mg ml⁻¹ in 0.02 M PBS pH 7.2) was added to 50 μ l of C $\bar{I}$ s (100 μ g ml⁻¹ in 10 mM Tris-HCl, 100 mM NaCl, 1 mM EDTA, pH 8.0) followed by 100 μ l of C $\bar{I}$ -Inh (present as either a 1:6 dilution of normal human serum or 1:250 dilution of purified C $\bar{I}$ -Inh (2,100 mg per 100 ml Immuno) in 0.02 M PBS pH 7.2). This commercially purified C $\bar{I}$ -Inh was used for all the assays illustrated in this text. Blank tubes were prepared identically by substituting 50 μ l of Tris buffer for the purified C $\bar{I}$ s. The volume of all tubes was brought to 500 μ l with phosphate buffer (100 mM sodium phosphate, 100 mM NaCl, 15 mM EDTA, pH 6.0) and incubated at 37°C for 15 min. Phosphate buffer (2 ml) was then added to all tubes. Immediately before assay, 50 μ l of substrate (3 mg ml⁻¹) was added to test and blank tubes which were then monitored at 340 nm, at room temperature, for 5 min.

(i) Purified C $\bar{I}$ s (ref. 10) was incubated with substrate, the reaction mixture monitored over 5 min and the change in absorbance recorded as ΔA .

(ii) C $\bar{I}$ s was incubated with C $\bar{I}$ -Inh followed by incubation with substrate. It can be seen from the table that the C $\bar{I}$ -Inh inhibits the conversion of substrate.

(iii) Patient or control IgG was incubated with C $\bar{I}$ s and C $\bar{I}$ -Inh. Patient but not control (control values in brackets) IgG prevented the C $\bar{I}$ -Inh from inactivating the C $\bar{I}$ s. The IgG used in this assay was isolated during a quiescent phase of the illness and was purified by ion exchange chromatography (IgG IEC)¹¹ or Protein A affinity chromatography (IgG Prot. A) or ion exchange chromatography followed by Sephadex G-200 gel filtration (IgG IEC:G200). (The latter two procedures were performed according to manufacturer's (Pharmacia) instructions.)

(iv) Pretreatment of patient IgG fractions with the protease inhibitor, soya bean trypsin inhibitor (SBTI) (Sigma type 1-S, 100 μ g ml⁻¹) had no effect on its activity, demonstrating that the ability of the IgG to impede the function of C $\bar{I}$ -Inh was not mediated by plasmin.

(v/vi) These readings demonstrate that patient IgG (IgG IEC) had no direct effect on C $\bar{I}$ s or on substrate.

hereditary form of angioedema. The finding of low serum C1q suggests activation of the classical complement pathway presumably mediated by antigen-antibody interaction. These findings, in this setting, are unique, as a clinical syndrome characterized by an acquired functional deficiency of C $\bar{I}$ -Inh has not previously been described.

It is recognized that in enzyme cascade systems, such as the complement system, the finding of a dysfunctional protein may be mediated by the presence of an autoantibody to that protein or its natural substrate^{4,5}. To determine if this patient's disorder could be mediated by an autoantibody to either C $\bar{I}$ or its inhibitor, IgG was isolated from his serum during active and quiescent phases of the illness. This IgG was assayed to determine if it could impede normal C $\bar{I}$ -Inh function (Table 1). All the patient IgG preparations prevented the inactivation of purified C $\bar{I}$ s by its natural inhibitor, C $\bar{I}$ -Inh. The phase of the illness from which the IgG was isolated did not affect its activity. IgG prepared from normal individuals using the same methods and adjusted to the same protein concentration had no effect on the activity of C $\bar{I}$ -Inh.

A variety of methods was used to isolate IgG, all based on different physical properties of this protein, that is, charge, affinity for protein A and molecular size. Using these different methods it was unlikely that the same contaminating serum proteins, which might have the capacity to interfere with the

assay, would be present in all the IgG preparations. For example, potential contaminants such as factor XII, kallikrein, plasmin or C $\bar{I}$ s have lower molecular mass than IgG and would therefore be separated from IgG by filtration with Sephadex G-200⁸. Furthermore, the IgG preparations from the patient and normal controls gave identical patterns when analysed by SDS-polyacrylamide gel electrophoresis indicating that there were no additional proteins in the patient's IgG.

Initial evaluation of these antibody preparations indicated that the inhibitory effect of the antibody could only be demonstrated if IgG was added before the interaction of C $\bar{I}$ -Inh with C $\bar{I}$ s. This may be explained by the irreversible nature of the C $\bar{I}$ s.C $\bar{I}$ -Inh interaction⁷. It was also observed that the effect of the antibody could be neutralized by the addition of increasing quantities of C $\bar{I}$ -Inh. If the quantity of C $\bar{I}$ -Inh was kept constant, higher concentrations of IgG completely prevented all C $\bar{I}$ -Inh function.

To demonstrate conclusively that the observed effect of the IgG preparation was due to the IgG molecule, the purified IgG was incubated with an anti-IgG (Fc) (Table 2). Precipitation of

Table 2 Effect of incubation of patients IgG with anti-IgG on C $\bar{I}$ -Inh function

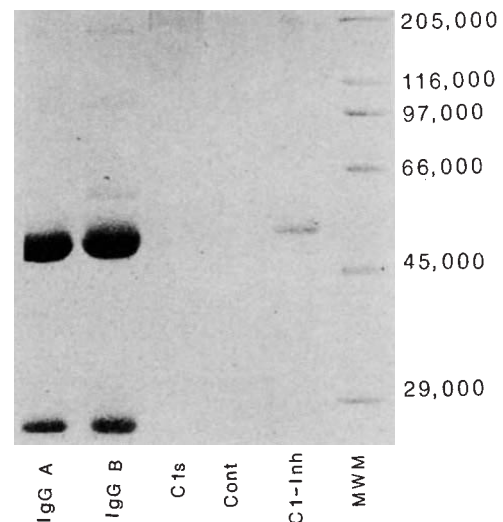
	ΔA
C $\bar{I}$ s + ZLNE	65
C $\bar{I}$ s + C $\bar{I}$ -Inh + ZLNE	1
C $\bar{I}$ s + 100 μ l (IgG 120 μ l + PBS 600 μ l) + C $\bar{I}$ -Inh + ZLNE	31
C $\bar{I}$ s + 100 μ l (IgG 120 μ l + anti-Fc 600 μ l) + C $\bar{I}$ -Inh + ZLNE	0
C $\bar{I}$ s + 100 μ l (PBS 120 μ l + anti-Fc 600 μ l) + C $\bar{I}$ -Inh + ZLNE	2
C $\bar{I}$ s + 100 μ l (IgG 150 μ l + PBS 100 μ l) + C $\bar{I}$ -Inh + ZLNE	48
C $\bar{I}$ s + 100 μ l (IgG 150 μ l + anti-FabA 100 μ l) + C $\bar{I}$ -Inh + ZLNE	28
C $\bar{I}$ s + 100 μ l (PBS 150 μ l + anti-FabA 100 μ l) + C $\bar{I}$ -Inh + ZLNE	2
C $\bar{I}$ s + 100 μ l (IgG 150 μ l + PBS 100 μ l) + C $\bar{I}$ -Inh + ZLNE	50
C $\bar{I}$ s + 100 μ l (IgG 150 μ l + anti-FabB 100 μ l) + C $\bar{I}$ -Inh + ZLNE	22
C $\bar{I}$ s + 100 μ l (PBS 150 μ l + anti-FabB 100 μ l) + C $\bar{I}$ -Inh + ZLNE	3

This assay used the same principle as outlined in Table 1. Purified C $\bar{I}$ s on addition to substrate resulted in a change in absorbance which in the presence of C $\bar{I}$ -Inh was greatly reduced. When patient IgG (IEC) was added, the activity of C $\bar{I}$ -Inh was impeded. The specific removal of IgG by the anti-IgG Fc-specific antisera (DAKO, rabbit IgG fraction) restored the activity of C $\bar{I}$ -Inh. With the addition of anti-Fab antisera (anti-FabA, monoclonal anti-human IgG Fab-specific (Miles-Yeda); anti-FabB, affinity-purified goat anti-human IgG Fab-specific (Sera-Lab)) the inhibitory effect of the IgG on C $\bar{I}$ -Inh was reduced by approximately 50%. Appropriate controls for each assay are shown. The experimental protocol was as follows: patient IgG (9 mg $\times$ ml⁻¹) and undiluted anti-IgG (Fc and anti-Fab) were incubated in the proportions shown at 37°C for 40 min. The samples were then centrifuged at 12,000g for 10 min to remove any precipitated IgG. Supernatant (100 μ l) was added to 50 μ l of C $\bar{I}$ s followed by 100 μ l of C $\bar{I}$ -Inh (1:250 in PBS). The assay then followed the procedure outlined in Table 1.

the IgG at antigen/antibody equivalence completely removed from the solution the capacity to inactivate C $\bar{I}$ -Inh. In contrast, in antibody excess (that is anti-IgG(Fc) excess), where no precipitation occurred, the activity of the IgG was not diminished. This demonstrated that the loss of activity of the IgG was not due to a nonspecific effect of the anti-IgG (data not shown). Having shown that the inhibitory effect resided in the IgG molecule further studies were performed to locate the active site. The IgG preparation was therefore incubated with monoclonal and polyclonal antisera to the Fab region in conditions which did not result in precipitation. As a result the activity of the IgG was diminished by 50%, indicating that the inhibitory effect was mediated by the Fab portion, and therefore probably by the antigen-binding site. Normal IgG treated in an identical fashion had no effect on C $\bar{I}$ -Inh function.

To determine if the activity of the patient's IgG was mediated through an effect on C $\bar{I}$ s or on C $\bar{I}$ -Inh, these purified proteins were coupled to Sepharose by cyanogen bromide and columns prepared for affinity chromatography (Table 3). The C $\bar{I}$ -Inh

Fig. 1 SDS-polyacrylamide gel electrophoresis of IgG preparations showing the following tracks; IgG A, patient IgG isolated by IEC; IgG B, control IgG isolated by IEC; CIs, material eluted from CIs column, no protein bands detected; Cont, material eluted from the control column, no protein bands detected; C1-Inh, material eluted from the C1-Inh column demonstrating two bands with molecular masses corresponding to those of the heavy and light chains of IgG. MWM, markers of relative molecular mass (M_r). The patient's clinical management was initially with the impeded androgen, danazol, which prevented episodes of angioedema and reverted all of the complement abnormalities, including the C1-Inh dysfunction. Unfortunately the patient subsequently became refractory to treatment with this agent. Prophylactic treatment with prednisolone (a standard treatment in autoimmune disorders), has now resulted in prevention of episodes of oedema in the patient. This finding is of interest because this agent is normally of no benefit in preventing attacks of angioedema¹². Two attempts to withdraw this drug resulted in recurrence of the oedema.



column completely removed the activity of the IgG whereas the CIs and control column had no effect. Pure IgG (Fig. 1) was eluted from the C1-Inh column but not from the CIs or control columns. The eluted IgG had the same inhibitory effect on C1-Inh function as the non-affinity purified IgG; however the specific activity of the affinity-purified preparation was approximately 50 times that of the non-affinity purified IgG. IgG prepared from normal individuals was not retained by any of the affinity columns.

Analysis by Western blotting of isoelectrically-focused, affinity-purified, patient IgG demonstrated that the antibody has the characteristic restricted mobility of a monoclonal antibody. The antibody reacted with anti- κ but not with anti- λ light chain antisera. Furthermore, in an experiment similar to that outlined

in Table 2, anti- κ antisera completely blocked the ability of the IgG to impede C1-Inh function. Anti- λ light chain antisera had no effect on the activity of the patient IgG. This antibody is therefore an IgG- κ antibody directed against the active site on C1-Inh. To date, despite continuing investigation the patient has no further evidence of lymphoproliferative disease.

The complement-fixing ability of patient IgG/C1-Inh complexes was assessed. IgG from patient and control was purified by ion exchange chromatography followed by gel filtration with Sephacryl S-300. This IgG, at concentrations of 3 and 6 mg ml⁻¹ was then incubated with normal serum at 37°C for 90 min. Following the incubation, CH₅₀ determinations were performed on the reaction mixture. Patient IgG did not cause significant reduction in CH₅₀, indicating that these immune complexes were not complement fixing. This is probably not surprising in view of the monoclonal nature of the antibody.

In summary, these studies demonstrated that this patient has an IgG antibody which binds to C1-Inh and prevents its function *in vitro*. Such an occurrence *in vivo* may adequately explain the major features of this case—the patient's clinical syndrome and dysfunctional C1-Inh.

It has recently been suggested that acquired angioedema is the consequence of idiotype-anti-idiotype interactions⁹. The mechanism by which such complexes activate complement and consume C1-Inh is not fully understood, nor is it appreciated why this does not happen in other diseases in which immune complexes are formed. The relationship of these anti-idiotype antibodies and the C1-Inh antibody found in this patient should be explored, as it is possible that they may be interrelated phenomena. These findings highlight the importance of autoantibodies in mediating disease. Further analysis of this antibody may help dissect the intricate role C1-Inh plays in the regulation of the inflammatory response.

Table 3 The effect of sepharose-linked CIs and C1-Inh on the activity of patient IgG

	ΔA
CIs + ZLNE	68
CIs + C1-Inh + ZLNE	2
CIs + IgG A + C1-Inh + ZLNE	34
CIs + IgG B + C1-Inh + ZLNE	31
CIs + IgG C + C1-Inh + ZLNE	3
CIs + el.A + C1-Inh + ZLNE	2
CIs + el.B + C1-Inh + ZLNE	1
CIs + el.C + C1-Inh + ZLNE	41

Sephacryl 4B (Pharmacia) was activated by cyanogen bromide and the purified proteins, CIs (ref. 10) 2 mg per ml of Sepharose and C1-Inh (Immuno) 8 mg per ml of sepharose, were coupled to the gels using a bicarbonate coupling buffer (NaHCO₃ 100 mM, NaCl 500 mM, pH 8.3). A control column with no protein attached was also prepared. Remaining uncoupled active sites were blocked using 100 mM Tris/HCl, pH 8. Columns of volume 3 ml were used. Equal volumes (1.5 ml) of ion exchange purified IgG (9 mg ml⁻¹) were incubated with the columns for 30 min at room temperature. This IgG was then washed through the columns with PBS (0.02 M pH 7.2) and used in the assay as described previously. These IgG preparations are termed IgG A (from the control column); IgG B (from the CIs column); and IgG C (from the C1-Inh column). When all the unbound IgG had been washed through the columns, any IgG which had bound to the Sepharose-linked proteins was eluted with diethylamine (diethylamine 50 mM, NaCl 150 mM, pH 11.2) and assayed as above (el.A, that eluted from the control column; el.B, from CIs column; el.C, from C1-Inh column). It can be seen from the table that the blocking effect of the IgG was completely removed by the C1-Inh column and not by the CIs or control columns. IgG was eluted from the C1-Inh column only and not from the CIs or control columns. This IgG (el.C) was shown to inactivate C1-Inh. IgG prepared from normal individuals was not absorbed by any of the affinity columns.

Received 10 March; accepted 29 July 1986

- Drachman, D. B. *New Engl. Med. J.* **298**, 136-142 (1978).
- Rose, N. R., Lorenzi, M. & Lewis, M. in *Basic and Clinical Immunology* (eds Stites, D. P. et al.) 652-655 (Lange, California, 1984).
- Petz, L. D. & Garratty, G. in *Acquired Immune Haemolytic Anaemia* (Churchill Livingstone, London, 1980).
- Lancet* **i**, 255 (1981).
- Lachmann, P. J. & Peters, D. K. in *Clinical Aspects of Immunology* (eds Lachmann, P. J. & Peters, D. K.) 38-40 (Blackwell, Oxford, 1982).
- Kaplan, A. P. et al. *CRC Crit. Rev. Immun.* **75**-93 (1981).
- Sim, R. B. & Reboul, A. *Meth. Enzym.* **80**, 43-54 (1981).
- Bloom, A. L. & Thomas, A. P. in *Haemostasis and Thrombosis* (Churchill Livingstone, London, 1981).
- Sheffer, A. L. et al. *J. Allerg. clin. Immun.* **75**, 640-646 (1985).
- Haines, A. L. & Lepow, I. H. *J. Immun.* **92**, 456-467 (1963).
- Fahey, J. L. & Terry, E. W. in *Handbook of Experimental Immunology* (ed. Weir, D. M.) 8.3-8.10 (Blackwell, Oxford, 1979).
- Frank, M. M., Gelfand, J. A. & Atkinson, J. P. *Ann. intern. Med.* **84**, 580-593 (1976).

Induction of CD4-dependent cell fusion by the HTLV-III/LAV envelope glycoprotein

Jeffrey D. Lifson*†‡, Mark B. Feinberg‡, Gregory R. Reyes§, Linda Rabin§, Babak Banapour§, Sekhar Chakrabarti||, Bernard Moss||, Flossie Wong-Staal‡, Kathelyn S. Steimer¶ & Edgar G. Engleman*†

* Department of Pathology, Stanford University, Stanford Medical School Blood Center, Palo Alto, California 94305, USA

† Department of Pathology, Palo Alto Veterans Administration Hospital, Palo Alto, California 94303, USA

‡ Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

§ Gene Labs, Inc., San Carlos, California 94070, USA

|| Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

¶ Chiron Corporation, Emeryville, California 94608, USA

* Present address: Gene Labs, Inc., San Carlos, California 94070, USA

Formation of syncytia, with progression to cell death, is a characteristic feature of *in vitro* cultures of susceptible cells infected with human T-lymphotropic virus type III/lymphadenopathy-associated virus (HTLV-III/LAV)¹⁻³. Viral antigen-positive multinucleated giant cells have also been observed in histological sections from infected individuals^{4,5}. *In vitro*, formation of these multinucleated giant cells occurs through cell fusion which is dependent on cell-surface expression of the differentiation antigen CD4 (ref. 1). Utilizing a recombinant vaccinia virus containing the gene for the envelope glycoprotein of HTLV-III/LAV⁶, we demonstrate that cell-surface expression of this protein, in the

absence of other HTLV-III/LAV structural or regulatory proteins, is sufficient to induce CD4-dependent cell fusion, leading to cell death, one of the characteristic manifestations of AIDS (acquired immune deficiency syndrome) virus cytopathology. This process may contribute to the loss of CD4⁺ T cells seen in AIDS.

Viral cytopathology, including cell fusion, has been associated with viral envelope glycoproteins⁷⁻¹². To determine whether HTLV-III/LAV-induced cell fusion could be attributed to properties of the viral envelope glycoprotein, we employed a recombinant vaccinia virus, containing the gene coding for the HTLV-III/LAV envelope glycoprotein. Cells infected with this recombinant virus (VSC-25) express the HTLV-III/LAV envelope glycoprotein, but do not synthesize any other HTLV-III/LAV structural or regulatory proteins⁶. A detailed description of this virus has been provided elsewhere⁶.

Infection of the CD4⁺ human T-cell line VB (ref. 1) with VSC-25 led to specific expression of the HTLV-III/LAV envelope glycoprotein, with immunoprecipitation demonstrating specific bands with relative molecular masses corresponding to those of appropriately processed envelope gene products (gp160, gp120, gp41) beginning 4-6 hours after infection^{13,14}. Immunofluorescence confirmed that cell surface expression of envelope glycoprotein was occurring (data not shown). Uninfected VB cells and VB cells infected with wild-type vaccinia virus or a control recombinant vaccinia virus (VSC-8) containing a non-envelope gene insert showed no evidence of HTLV-III/LAV envelope glycoprotein expression.

Multinucleated giant cells were observed 4-6 hours after infection with VSC-25, which correlates with the kinetics of HTLV-III/LAV envelope glycoprotein expression. The syncytia were morphologically indistinguishable from those seen in cultures infected with HTLV-III/LAV. The multinucleated giant cells underwent rapid, progressive degeneration, with cytoplasmic ballooning, nuclear pyknosis, and eventual extrusion of the condensed nuclear mass¹. Virtually all the multinucleated cells were non-viable, as assessed by Trypan blue exclusion, within

Fig. 1 The cell-surface expression of HTLV-III/LAV envelope glycoprotein induces cell fusion. For infection, H9 cells were incubated for 30 min at 37 °C with stock virus preparations of wild type vaccinia virus, VSC-8 or VSC-25, diluted in phosphate buffered saline to give a multiplicity of infection of 40. After washing, infected H9 cells were labelled with rhodamine isothiocyanate (RITC; red) and uninfected CD4⁺ VB cells were labelled with fluorescein isothiocyanate (FITC; green) prior to co-culture. After 5 h, cells were fixed for fluorescence microscopy³⁰. Fig. 1a-c RITC-labelled H9 cells infected with VSC-25, photographed to reveal RITC (a) and FITC (b) fluorescence and under phase contrast (c). Panels d-i show uninfected FITC-labelled CD4⁺ VB cells, photographed to show RITC (d) and FITC (e) fluorescence, and under phase contrast (f). RITC-labelled H9 cells infected with VSC-25 and cultured with FITC-labelled CD4⁺ VB cells gave rise to syncytia showing both red and green fluorescence, indicative of cell fusion (g-i). Small cells surrounding giant cells show only green, or only red fluorescence, while giant cells show bright fluorescence for RITC and FITC. Original magnification, $\times 400$.

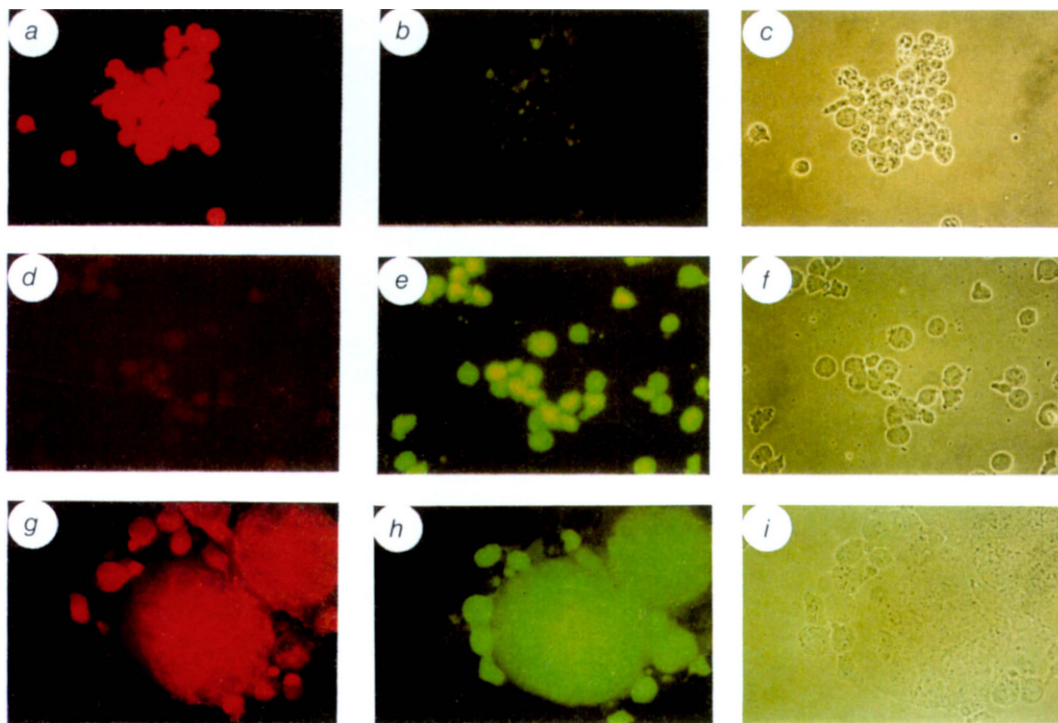


Table 1 Serum neutralization of HTLV-III/LAV infectivity and envelope-induced cell fusion

Serum designation	HTLV-III/LAV antibody status	Neutralization titre for free virus infectivity	Neutralization of cell fusion: H9 _{HTLV-III} + VB scored at:						Neutralization of VSC-25 induced VB cell fusion scored at:					
			3 hours			6 hours			6 hours			8 hours		
			1:10	1:50	1:100	1:10	1:50	1:100	1:20	1:40	1:80	1:20	1:40	1:80
30001	—	0	3+	3+	3+	4+	4+	4+	2+	2+	2+	4+	4+	4+
30003	—	0	3+	3+	3+	4+	4+	4+	2+	2+	2+	4+	4+	4+
20025	+	12,000	—	—	3+	—	4+	4+	—	—	—	—	—	1+
20006	+	5,000	—	—	3+	2+	3+	4+	—	—	—	—	—	—
20019	+	1,700	—	1+	3+	4+	4+	4+	—	1+	1+	—	1+	2+
20012	+	1,000	2+	3+	3+	4+	4+	4+	—	—	1+	—	—	2+
20030	+	700	1+	3+	3+	4+	4+	4+	—	1+	1+	1+	2+	3+
20059	+	180	2+	3+	3+	4+	4+	4+	—	1+	2+	2+	2+	4+

Seropositive samples were from individuals scoring positive in a licensed HTLV-III/LAV antibody enzyme-linked immunosorbent assay (ELISA), with specific reactivity confirmed by immunoblot. Seronegative specimens were from randomly chosen healthy blood donors. All sera were heat inactivated (56°C for 1 h) prior to assay. For assay of neutralization of free virus infectivity, serum samples diluted in RPMI-1640 with 10% fetal calf serum were mixed and incubated for 30 min at room temperature with an equal volume of AIDS virus-containing supernatant from a culture of HUT-78 cells infected with ARV-2 (ref. 29). Fresh HUT-78 cell cultures in 24-well plates (1×10^4 cells per ml per well) were inoculated with 100 μ l of serum/virus mixture, starting at a serum dilution of 1:20, with serial threefold dilutions thereafter. After 7 days culture, cells were collected and lysed in 1 ml of PBS containing 1% (v/v) Triton X-100. Infection was monitored by determining the amount of viral core antigen (p24^{gag}) with an ELISA system²⁹ which incorporates a murine anti-p24 monoclonal antibody and a high titre p24 reactive rabbit antiserum. Neutralizing titres reported are the reciprocals of the dilutions which caused 50% inhibition of p24 production in infected cultures. To measure inhibition of fusion induced by HTLV-III/LAV infected cells, 5×10^4 HTLV-III_g infected H9 cells were mixed with 5×10^4 CD4⁺ VB cells in a total volume of 100 μ l in the presence of the indicated dilutions of test serum in the wells of 96-well microtitre plates. Each dilution was tested in triplicate. After culture at 37°C, syncytium (defined as ≥ 4 nuclei within a common cell membrane) formation was scored by inverted microscope ($\times 300$) at the indicated time points as: (—) no syncytia; (1+) rare small syncytia; (2+) multiple moderately sized syncytia; (3+) large syncytia in most, but not all fields examined; (4+) numerous large syncytia in all fields examined. To measure serum inhibition of HTLV-III/LAV envelope glycoprotein induced cell fusion, CD4⁺ VB cells were infected with VSC-25, then incubated in the presence of the indicated dilutions of test serum. Syncytium formation was scored as described. In separate experiments, sera from an additional five HTLV-III/LAV seronegative donors did not neutralize HTLV-III/LAV or VSC-25-induced cell fusion. Sera from two additional sources demonstrated to have neutralizing activity in two separate assays of neutralization of HTLV-III virion infectivity similarly inhibited cell fusion (data not shown).

24 to 36 hours of infection. Parallel cultures infected with wild-type vaccinia virus or VSC-8 did not exhibit such changes.

To further study syncytium induction by VSC-25, we used a two-colour fluorochrome surface labelling procedure¹. H9 cells² were infected with either wild-type vaccinia virus, VSC-8, or VSC-25, then washed and surface labelled with rhodamine isothiocyanate (RITC)¹. Uninfected CD4⁺ VB cells were labelled with fluorescein isothiocyanate (FITC)¹. The labelled H9 and VB cells were then washed, mixed in various combinations, and observed for development of multinucleated giant cells. Formation of syncytia resulting from fusion between infected H9 cells and uninfected VB cells was demonstrated by simultaneous RITC and FITC fluorescence of multinucleated giant cells (Fig. 1). These cells were observed only in cultures containing H9 cells infected with VSC-25 and not in uninfected H9 cultures or in H9 cells infected with wild-type vaccinia virus, or VSC-8. Syncytia were observed 4–6 hours after mixing of the two populations. In the absence of added VB cells, H9 cells underwent limited fusion when infected with VSC-25; this process occurred more slowly. The extent and kinetics of HTLV-III/LAV envelope glycoprotein-induced cell fusion parallel results obtained with H9 and VB cells infected with HTLV-III/LAV, and correlate with the relative amounts of CD4 expressed on the surface of the cells (B. S. Stein, J.D.L., K. G. Bensch and E.G.E., manuscript in preparation).

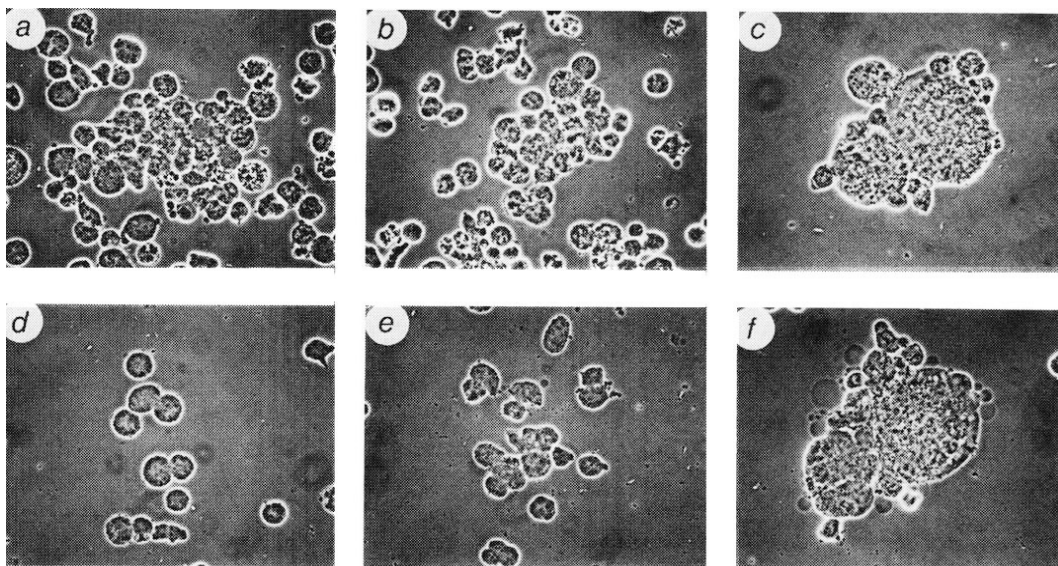
To confirm the role of CD4 in VSC-25-induced fusion, cells from the CD4⁺ VB cell line and a CD4⁺ VB variant subline¹ were infected with wild type vaccinia virus, VSC-8 or VSC-25 and observed for formation of multinucleated giant cells. Only cells expressing surface CD4 formed syncytia after infection with the recombinant vaccinia virus VSC-25. This process was specifically blocked by a monoclonal antibody to CD4, but not by antibodies to other surface antigens (CD3, CD5, CD7; data not shown). Similar results were obtained when uninfected CD4⁺ and CD4⁺ VB cells were mixed with VSC-25 infected H9 cells: the CD4⁺ VB cells did not fuse with infected H9 cells, whilst the CD4⁺ VB cells fused in a manner specifically blocked by anti-CD4 antibody (Fig. 2). CD4⁺ VB cells infected with VSC-25 did not undergo fusion, but fused readily with uninfected CD4⁺ VB cells.

In additional experiments, CD4⁺ or CD4⁺ cells of T and B lymphoid or non-lymphoid lineage were infected with VSC-25, then surface labelled with RITC as described above. The labelled cells were then cultured alone, or co-cultured with FITC-labelled CD4⁺ VB cells, and observed for evidence of cell fusion. In contrast to CD4⁺ T cells, cells from the CD4⁺ Epstein-Barr virus (EBV)-transformed B lymphoblastoid cell line Raji, the CD4⁺ erythroid/myeloid leukaemia cell line K562 and the CD4⁺ rat sarcoma cell line X/C, and CD4⁺ T cells CD4⁺ T cells (HSB cell line) did not undergo fusion after infection with VSC-25. Cells of the EBV-transformed B cell line LCo-B, which express low levels of CD4, underwent limited fusion. Extensive fusion was noted for all these cell lines when VSC-25 infected cells were co-cultured with uninfected CD4⁺ VB cells. HTLV-III/LAV envelope glycoprotein-induced cell fusion is thus CD4 dependent, but does not appear to be restricted by lineage, species, or other host cell factors.

Finally, we determined the ability of known HTLV-III/LAV antibody positive sera to inhibit envelope glycoprotein-induced cell fusion, and compared these results with the ability of the sera to inhibit infectivity of free HTLV-III/LAV virions^{15,16}. As shown in Table 1, a number of the sera did inhibit AIDS virus-induced cell fusion, and the presence of this inhibitory activity generally correlated with the ability of the sera to inhibit infectivity of HTLV-III/LAV virions. Inhibition of syncytium induction could only be demonstrated at low titre, even for sera which had relatively high neutralizing activity for cell-free virus infectivity. Most importantly, the ability of the sera to inhibit fusion between HTLV-III/LAV infected H9 cells and uninfected CD4⁺ VB cells correlated with their ability to inhibit fusion of VB cells infected with VSC-25, suggesting that the neutralizing activity present in the sera is directed against envelope determinants.

AIDS virus-induced cell fusion thus involves specific interactions between cell surface displayed envelope glycoprotein of HTLV-III/LAV and CD4. In the recombinant vaccinia virus system used in these studies, the HTLV-III/LAV envelope glycoprotein is expressed independently of the influence of other AIDS virus structural or regulatory proteins. Envelope glycoprotein expressed on the cell surface under these conditions is

Fig. 2 CD4-dependence of HTLV-III/LAV envelope glycoprotein-induced cell fusion. Cells were infected as described in the legend to Fig. 1. Uninfected H9 cells (a) and VB cells (b) did not show spontaneous fusion. VSC-25 infected H9 cells mixed with CD4⁺ VB cells showed extensive fusion (c) while H9 cells infected with wild-type vaccinia virus or VSC-8 did not fuse with CD4⁺ VB cells (d). VSC-25 infected H9 cells did not fuse with a CD4⁻ VB cell variant¹ (data not shown). Monoclonal antibody to CD4 (antibody S3.5 (ref. 1) (10 $\mu\text{g ml}^{-1}$) blocked the fusion of VSC-25 infected



H9 cells and CD4⁺ VB cells (e) while antibodies to CD3, CD5 and CD7, at the same concentration, did not block fusion. f Shows representative results for anti-CD7. Original magnification, $\times 500$.

sufficient to induce CD4-dependent cell fusion, leading to death of the syncytial cells. Other HTLV-III/LAV proteins are not directly involved in the fusion process. In particular, the *tat* gene¹⁷⁻¹⁹ is not present in VSC-25, nor is the recently described *art* gene²⁰. While *tat* is thus not directly required for cell fusion, it may enhance the fusion process by increasing the density of envelope glycoprotein on the surface of infected cells¹⁷⁻¹⁹. Although VSC-25 contains the coding sequence for the 3'-*orf* gene of HTLV-III/LAV, the product of this gene is not expressed by VSC-25-infected H9 cells (M.B.F., unpublished observations), and deletion mutant studies have demonstrated that the 3'-*orf* gene is not required for HTLV-III/LAV-induced cytopathology^{21,22}.

Fusion leads to death of the syncytial cells. While HTLV-III/LAV may also cause cell death through mechanisms independent of cell fusion, our results suggest that this process may contribute to the progressive depletion of CD4⁺ T cells observed in patients with AIDS. The presence of viral antigen positive syncytia in tissue sections suggests that HTLV-III/LAV-induced cell fusion may occur *in vivo*. In our *in vitro* model, a single infected cell displaying viral envelope glycoprotein on its surface can recruit many uninfected CD4⁺ cells to a syncytium which subsequently dies, a finding of potential pathogenic significance.

Envelope glycoprotein-CD4 interactions are also involved in the infection of CD4⁺ cells by cell-free HTLV-III/LAV virions²³⁻²⁵. It is not known whether the epitopes involved in the two distinct processes are the same, but the observed correlation between the ability of patient sera to block *in vitro* infectivity of cell-free virions and inhibit envelope glycoprotein-induced syncytium formation is consistent with the two processes being mediated through similar mechanisms. In addition, the demonstration that sera from HTLV-III/LAV infected individuals are capable of inhibiting these envelope glycoprotein dependent-processes has important implications for the study of immune responses to the AIDS virus, and the possibility of developing a vaccine. The observation that sera which neutralized the infectivity of ARV-2 virions also neutralized cell fusion mediated by HTLV-III_b-infected cells, despite the documented genomic heterogeneity in the envelope sequences of these distinct AIDS virus isolates^{26,27}, suggests that functionally significant, conserved regions of the envelope are immunogenic *in vivo*. The limited capacity of the sera to neutralize cell fusion, particularly at later time points in the fusion assays, may reflect the cumulative synthesis of sufficient envelope antigen to satur-

ate the amount of neutralizing antibody present. Alternatively, the interaction of neutralizing antibodies with viral antigens may be of lower affinity than interactions between viral antigens and host cells, as was recently described for visna virus²⁸. This might explain the persistence of virus, despite the presence of neutralizing serum antibodies, and would explain the kinetics of neutralization of cell fusion, the relatively low neutralizing titre of most sera, and the observation that the presence of neutralizing serum antibody does not appear to confer protection against the development of HTLV-III/LAV-related disease. While the present studies raise concern regarding the potential use of recombinant viruses such as VSC-25 as prophylactic vaccines, the lower density of CD4 on peripheral blood T cells compared to the VB cells used in our assay, and effective anti-vaccinia responses in immunocompetent individuals might limit the extent of cell fusion induced by such a recombinant virus *in vivo*. Finally, these studies emphasize the usefulness of the recombinant vaccinia virus system for studying the biological properties of individual proteins of other viruses.

We thank Dr J. C. Chermann for providing the LCo-B cell line, Dr S. Smith for the VB cell line, Drs M. Robert-Guroff and D. Bolognesi for characterized patient sera, Dr P. Earl for preparation of virus stocks, Dr O. Narayan for providing data prior to publication, Dr J. Levy for ARV-2 virus, Drs M. McGrath, B. Stein, R. Jarrett and C. Benike for review of the manuscript, P. Verzola and P. Horne for photographic assistance, A. Keylor for preparation of figures, and D. Jones for preparation of the manuscript. This work was supported in part by NIH grant HL33811, and by the Medical Research Service of the Veterans Administration. J.D.L. is the recipient of a Career Development Award from the Veterans Administration. M.B.F. is a predoctoral fellow in the Cancer Biology Program, Stanford University.

Note added in proof: Since submission of this manuscript observations indicating the involvement of the HTLV-III/LAV envelope glycoprotein in virally induced cell fusion have also been reported by Sodroski *et al.* (*Nature* **322**, 470-474 (1986)).

Received 19 June 1986; accepted 19 August 1986.

1. Lifson, J. D., Reyes, G. R., McGrath, M. S., Stein, B. S. & Engleman, E. G. *Science*, **232**, 1123-1127 (1986).
2. Popovic, M., Sarnadharan, M. G., Read, E. & Gallo, R. C. *Science*, **224**, 497-500 (1984).
3. Barré-Sinoussi, F. *et al.* *Science* **220**, 868-872 (1983).
4. Sharer, L. R., Eun-Sook, C. & Epstein, L. G. *Hum. Path.* **16**, 760-761 (1985).
5. Barnes, D. M. *Science* **232**, 1091-1093 (1986).

6. Chakrabarti, S., Robert-Guroff, M., Wong-Staal, F., Gallo, R. C. & Moss, B. *Nature* **320**, 535-537 (1986).
7. Dörner, A. J. & Coffin, J. M. *Cell* **45**, 365-374 (1986).
8. Merz, D., Scheid, A. & Choppin, P. W. *J. Exp. Med.* **151**, 275-288 (1980).
9. Pinter, A., Chen, T., Lowey, A., Cortez, N. G. & Silagi, S. *J. Virol.* **57**, 1048-1053 (1986).
10. Nagy, K., Clapham, P., Cheingsong-Popov, R. & Weiss, R. A. *Int. J. Cancer* **32**, 321-328 (1983).
11. Chatterjee, S. & Hunter, E. *Virology* **107**, 100-108 (1970).
12. White, J., Kielian, M. & Helenius, A. *Q. Rev. Biophys.* **16**, 151-195 (1983).
13. Allan, J. S. *et al. Science* **228**, 1091-1093 (1985).
14. Barin, F. *et al. Science* **228**, 1094-1096 (1985).
15. Robert-Guroff, M., Brown, M. & Gallo, R. C. *Nature* **316**, 72-74 (1985).
16. Weiss, R. A. *et al. Nature* **316**, 69-71 (1985).
17. Fisher, A. G. *et al. Nature* **320**, 367-371 (1986).
18. Sodroski, J. *et al. Science* **227**, 171-173 (1985).
19. Arya, S. K., Guo, C., Josephs, S. & Wong-Staal, F. *Science* **229**, 69-73 (1985).
20. Sodroski, J. *et al. Nature* **321**, 412-417 (1986).
21. Fisher, A. G. *et al. Science* **233**, 655-659 (1986).
22. Sodroski, J. *et al. Science* **231**, 1549-1552 (1986).
23. Dalgleish, A. G. *et al. Nature* **312**, 763-767 (1984).
24. Klatzmann, D. *et al. Nature* **312**, 767-769 (1984).
25. McDougal, J. S. *et al. Science* **231**, 382-385 (1986).
26. Sanchez-Pescador, R. *et al. Science* **227**, 484-492 (1985).
27. Ratner, L. *et al. Nature* **312**, 277-284 (1985).
28. Kennedy-Stoskopf, S. & Narayan, O. *J. Virol.* **59**, 37-44 (1986).
29. Steimer, K. *et al. Virology* **150**, 283-290 (1986).
30. Lifson, J. D., Sasaki, D. T. & Engleman, E. G. *J. Immun. Meth.* **86**, 143-150 (1986).

Isolation and expression of a complementary DNA that confers multidrug resistance

Philippe Gros*, Yinon Ben Neriah†, James M. Croop‡ & David E. Housman‡

* Department of Biochemistry, Faculty of Medicine, McGill University, Montreal, Canada H36 1Y6

† Whitehead Institute and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

‡ Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The emergence and outgrowth of a population of tumour cells resistant to multiple drugs is a major problem in the chemotherapeutic treatment of cancer. We have used highly drug-resistant cell lines developed *in vitro* to study the molecular basis of multidrug resistance. In these cell lines high levels of resistance are frequently associated with amplification and overexpression of a small group of genes termed *mdr* (refs 1-3) or *gp170* (ref. 4). Direct evaluation of the role of these genes in multidrug resistance has awaited the isolation of a member of this gene family in a biologically active form. Here we report the isolation of DNA clones complementary to the cellular messenger RNA transcripts of *mdr* genes and show that high-level expression of a full-length complementary DNA clone in an otherwise drug-sensitive cell confers a complete multidrug-resistant phenotype. Our results demonstrate that overexpression of a single member of the *mdr* group is sufficient to confer drug resistance. Furthermore, because the cDNA was isolated from a drug-sensitive cell, mutations in the primary sequence of *mdr* are not required to produce a multidrug-resistance phenotype.

We used the technique of in-gel renaturation⁵ to identify, clone and characterize¹ a genomic DNA segment (~140 kilobases, kb) commonly amplified⁶ in two independently derived multidrug-resistant hamster cell lines, C5 (ref. 7) and LZ (ref. 8). This segment contains a gene of ~75 kb in size that appears to be a member of a small group of related genes termed *mdr*. The cloned hamster *mdr* gene encodes a ~5-kb mRNA whose level of expression is correlated with the level of resistance of various multidrug-resistant hamster lines¹. The human and mouse homologues of the cloned *mdr* gene have also been found amplified and overexpressed in multidrug-resistant human and mouse cell lines^{2,3,9}.

To isolate cDNA clones, a size-selected cDNA library was constructed in phage λ gt11 from mRNA extracted from a drug-

sensitive mouse pre-B-cell line. Two genomic probes, pDR7.8 and pDR1.6, derived from the cloned *mdr* region amplified in multidrug-resistant LZ hamster cells¹, were used to screen the cDNA library. These two subclones were chosen because they are known to hybridize to the hamster *mdr* mRNA(s) overexpressed in LZ cells. Screening of 500,000 size-selected cDNA clones yields 42 positives which fall into two groups according to their *Eco*RI and *Bgl*III restriction maps. The two longest representative members of the groups, λ DR29 (3.9 kb) and λ DR11 (4.3 kb), are shown in Fig. 1A. Individual clones belonging to these two groups were detected at equal frequency in two rounds of screening of the cDNA library, suggesting that mRNAs corresponding to each class of cDNA are transcribed at equivalent levels in drug-sensitive cells.

Preliminary DNA sequence analysis from three overlapping cDNA clones of the group represented by λ DR11 suggested that this clone is likely to contain a complete coding sequence for a polypeptide. A poly(A) segment was identified at one end, whereas at the opposite end, we observed a long continuous open reading frame preceded by a 120 base pair (bp) segment that included termination codons in all three reading frames. Within the first 250 bp of open reading frame, 5 AUG codons were identified (data not shown). Hybridization studies show that λ DR11 is homologous with segments of the entire genomic domain cloned from LZ hamster cells (Fig. 1B). Subcloned fragments of λ DR11 hybridize strongly to 8 cosmid clones overlapping the hamster *mdr* domain. Thus, the hamster *mdr* gene is between 50 and 75 kb in size and transcription proceeds from left to right within that gene.

The biological activity of the cDNA insert of λ DR11 was tested in transfection experiments. The 4.3-kb *Eco*RI insert of λ DR11 was introduced in the eukaryotic expression vector p91023B in the sense and anti-sense orientations to generate plasmids pDREX4 and pDREX5, respectively. The basic features of this expression system have been described elsewhere^{10,11}. LR73 (ref. 12) drug-sensitive hamster cells were transfected with plasmids pDREX4 or pDREX5 (Fig. 2A) and their ability to survive drug selection tested using the chemotherapeutic agent Adriamycin (ADM) at 0.05 and 0.1 μ g ml⁻¹. Transfection of pDREX4 DNA into LR73 cells results in a 200-fold increase in the number of colonies resistant to 0.05 μ g ml⁻¹ Adriamycin (Fig. 2A,a,b) over background colony numbers observed in dishes from control LR73 cells transfected with pDREX5 (Figure 2A,b), p91023B or salmon-sperm (SS) DNA (data not shown). At the higher dose of Adriamycin (0.1 μ g ml⁻¹) pDREX4 behaved as a dominant selectable marker: although no colonies were observed in control groups transfected with pDREX5 (Fig. 2A,c) p91023B or SS DNA (data not shown) an average of 75-150 colonies emerged on dishes of LR73 cells transfected with pDREX4 (transfection efficiency of 200 colony-forming units (CFU) per 10⁶ cells per μ g DNA). Identical results were obtained in a parallel experiment where colchicine was used as a cytotoxic drug at 0.1 and 0.3 μ g ml⁻¹ (data not shown). To examine further the role of the *mdr* cDNA in the pleiotropic drug-resistance phenomenon, four individual transfected clones were examined for their level of cross-resistance to ADM, colchicine (COL) and vincristine (VCR) (Fig. 2B). Results show that the 4 clones tested are cross-resistant to ADM, COL and VCR: a 10-20-fold increase in resistance against the three drugs tested above the LR73 control was observed (see Fig. 2B legend).

To ascertain that amplification of the endogenous copy of the hamster gene does not contribute to the resistance phenotype of the transfectants, genomic DNA from individual transfected clones was digested with *Eco*RI and analysed by Southern blotting. Duplicate blots were hybridized to either the cDNA insert (Fig. 3a) or the vector probes (Fig. 3b). When the cDNA fragment is used as hybridization probe it detects three *Eco*RI fragments present at single-copy level in LR73 cells and in each transfected clone (Fig. 3a) representing the endogenous non-

† Present address: Department of Immunology, The Hebrew University, Hadassah Medical School, Jerusalem, 91010 Israel.

Fig. 1 A, Schematic representation of two cDNA inserts representative of the two independent populations of clones isolated from the mouse cDNA library. The cDNA inserts from the λ DR29 group contain two internal *EcoRI* sites as digestion of several clones yield constant 1.1- and 0.9-kb fragments with a third fragment of variable size. The second group, represented by clone λ DR11, does not show internal *EcoRI* sites. Restriction analysis with enzymes *Bgl*II and *EcoRI* in combination yields cDNA fragments of 1.5 and 1.3 kb with a third fragment of variable size. The 1.5- and 1.4-kb *Bgl*II/*EcoRI* fragments as well as the 1.3-kb internal *Bgl*II fragment of cDNA insert from phage λ DR11 were subcloned in puc19 (ref. 18) to generate subclones pcDR 1.5, pcDR 1.4 and pcDR 1.3. Scale bar, 1 kb. B, Mapping of the cloned cDNA to the amplified unit cloned in cosmid vectors from multidrug-resistant LZ hamster cells¹. Cosmid DNA from clones cosDR1B, 6A, 3A, 7K, 12X, 5X, 12W and 4W were digested with *Bam*HI, separated by electrophoresis and the blots hybridized to ³²P-labelled¹⁹ DNA inserts from clones pcDR1.4 (a), pcDR1.3 (b) and pcDR1.5 (c). Arrows, position and size (in kb) of hybridizing internal *Bam*HI fragments from overlapping cosmid clones. Other hybridizing fragments homologous hybrid fragments containing the cosmid vector pSAE²⁰ (~9 kb). **Methods.** A cDNA library was constructed using cytoplasmic RNA from the BALB/c mouse line 70Z/3 (ref. 21) as previously described², with the following modifications. After addition of *EcoRI* linkers to the double-stranded cDNA mixture, the DNA was electrophoresed on a 1% low melting agarose gel (LMP agarose, BRL). The portion of the gel corresponding to DNA of size >2.3 kb was cut out and the DNA was recovered by phenol and chloroform extractions and ethanol precipitation. Aliquots of cDNA (10 ng) were ligated to 300 ng *EcoRI*-digested λ gt11 cloning arms, packaged into phage particles and the library was plated onto *Escherichia coli* Y1088. The plated library was screened with pDR1.6 and pDR7.8 by plaque hybridization³.

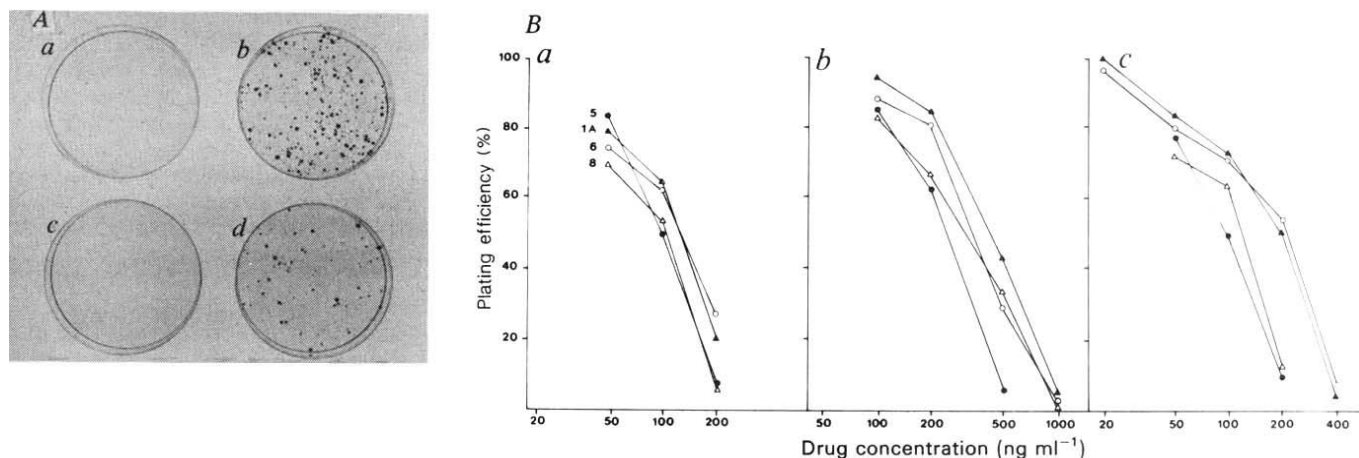
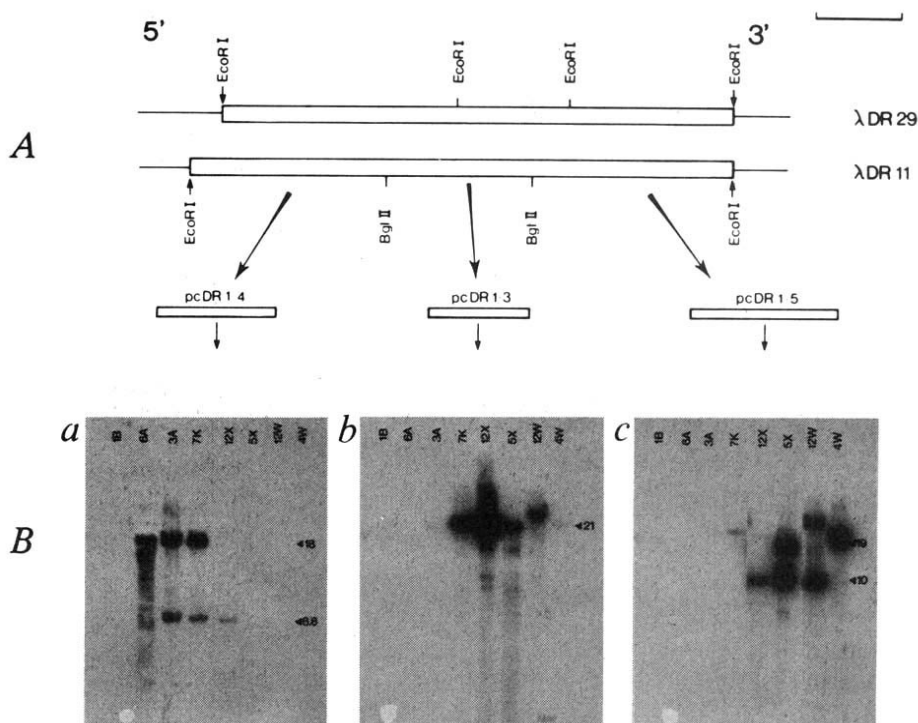
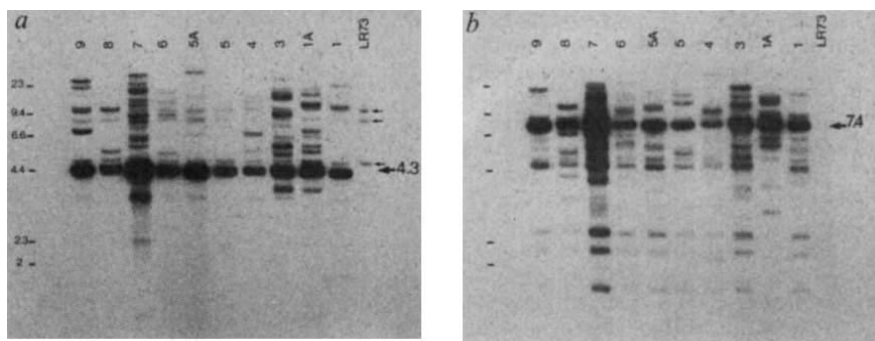


Fig. 2 A, Transfection of cDNA clones into drug-sensitive hamster cells. DNA transfection of LR73 hamster cells¹² was performed essentially as described elsewhere²⁴. LR73 cells were exposed to a calcium phosphate precipitate of plasmid pDREX5 (a, c) or plasmid pDREX4 (b, d). Forty-eight hours later, the cells were plated in ADM at 0.05 μ g ml⁻¹ (a, b) or 0.1 μ g ml⁻¹ (c, d). Five weeks later the colonies were fixed, 4 h with 4% formaldehyde, and stained with 0.1% methylene blue. B, Characterization of the drug-resistance phenotype of individual transfectants. Four independent transfectants selected in 0.1 μ g ml⁻¹ ADM were tested for cross-resistance to other cytotoxic drugs. Cells (500) of clone 6, 1A, 5, 8 and control LR73 cells were plated in medium containing a, ADM (0, 0.05, 0.1, 0.2, 0.5 and 1 μ g ml⁻¹); b, colchicine (0, 0.1, 0.2, 0.5, 1, 2 μ g ml⁻¹); and c, vincristine (0, 0.02, 0.04, 0.1, 0.2 and 0.4 μ g ml⁻¹). At day 14 colonies were fixed, stained and counted. Results are presented as the relative plating efficiency of each clone expressed as a percentage of the plating efficiency of individual control cells grown in absence of drugs versus the logarithm of the drug concentration. D10 is the drug dose necessary to produce a 10% plating efficiency. The D10 value for LR73 controls in the three drugs tested was ≤ 20 ng ml⁻¹ and is not presented. The D10 values for the 4 clones tested was ≥ 200 ng ml⁻¹ for ADM, VCR and ≥ 500 ng ml⁻¹ for COL.

amplified copy of the hamster *mdr* gene. It also detects an intensely hybridizing 4.3-kb band representing the intact cDNA *EcoRI* fragment of pDREX4 present at copy number varying between 10 (clone 5) and 50–60 copies (clone 7) per genome. In addition to this 4.3-kb band, the cDNA probe detects several *EcoRI* fragments >4.3 kb which may originate from integrated

copies of pDREX4 where integration has resulted in rearrangement of the 4.3-kb cDNA fragment. When the cloning vector p91023B is used as hybridization probe on this blot, a similar pattern emerges (Fig. 3b). These results demonstrate that the multidrug-resistance phenotype of the clones tested is linked to the presence of multiple copies of pDREX4, suggesting that

Fig. 3 Southern analysis of genomic DNA of cells transfected with pDREX4. Genomic DNA (10 µg) from clones 1, 1A, 3, 4, 5, 5A, 6, 7 and 8 selected in 0.1 µg ml⁻¹ ADM and from the hamster LR73 line were digested with *Eco*RI. Digestion products were electrophoresed in a 1% agarose gel containing 0.04 M Tris acetate (pH 8), 0.002 M EDTA, transferred to a hybridization membrane (Zetaprobe, Biorad) and the blot was hybridized to 10⁷ c.p.m. ³²P-labelled insert DNA from (a) pcDR1.2 or (b) to the expression vector p91023B. Small arrows in a, position of the hybridizing fragments corresponding to the endogenous copy of the hamster gene homologous with the probe; large arrow in a, position of the intact 4.3-kb *Eco*RI cDNA insert of pDREX4 and in b shows the 7.4-kb p91023B *Eco*RI fragment of pDREX4. Size markers, *Hind*III-digestion products of phage λ.



survival in the test concentration of ADM requires the expression of multiple copies of pDREX4.

The cDNA molecules introduced in the *Eco*RI site of p91023B are transcribed into a hybrid mRNA containing the cDNA sequence flanked by the adenovirus tripartite leader at the 5' side and by the mouse dihydrofolate reductase (*dhfr*) sequence serving as a non-translated sequence at the 3' side¹². This situation is advantageous as it also allows a distinction between the endogenous *mdr* mRNA transcribed from the hamster gene and the mRNA transcribed from the transfected cDNA construct (~1.1 kb larger). When a cDNA subclone (pcDR1.3) was used as a hybridization probe to screen Northern blots prepared from RNA isolated from individual drug-resistant clones, it detected a ~6-kb band. The shorter endogenous *mdr* cellular transcript (~5 kb) is not detected on this autoradiograph (2 h exposure, Fig. 4a) but can be visualized only after a much longer exposure of the blot (48 h exposure, Fig. 4b). Hybridization of the blot to a mouse *dhfr* probe supports the contention that the 6-kb band is a hybrid mRNA containing both the *mdr* and the *dhfr* sequences (data not shown). High levels of *mdr* mRNA are present in all drug-resistant transfectants tested compared with LR73 control cells (Fig. 4a). A dot-blot analysis (Fig. 4c) shows that all transfectants present a minimum of an 8–16-fold increase in *mdr* cDNA expression compared with the background level of *mdr* RNA of LR73 (not detectable at this exposure time).

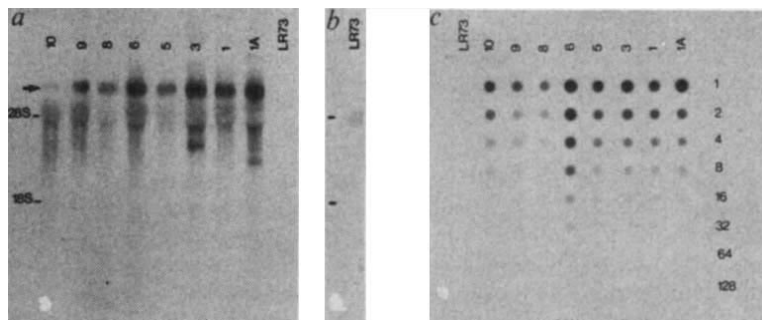
In summary, the pleiotropic drug-resistance phenotype can be conferred to drug-sensitive cells by transfection with the *mdr* cDNA constructed in pDREX4. Our experimental results indicate that in independent drug-resistant clones, multiple copies of the transfected plasmid are associated with high levels of expression of *mdr* cDNA. These results establish a direct causal relationship between the expression of the cloned *mdr* cDNA and the multidrug resistance phenotype.

The experimental results obtained from our transfection studies with clone pDREX4 answer two additional questions. First, although amplification of two *mdr* gene homologous with cDNA clones λDR29 and λDR11 has been detected in highly

resistant cultured cells (ref. 1; P.G. and M. Raymond, unpublished), our results indicate that overexpression of a single member of the *mdr* group is sufficient to produce a multidrug-resistant phenotype *in vitro*. Although the role and function of other members of the *mdr* group await further characterization of cDNA clones, amplification/overexpression of more than one *mdr* gene may be necessary to achieve the very high levels of resistance detected in cell lines such as LZ and C5. Second, it has been suggested that somatic mutations in the structural *mdr* gene may be involved in the establishment of resistance in cultured cells^{9,13}. However, we show here that transfection with a cDNA isolated from a drug-sensitive cell line can confer the complete resistance phenotype to these cells. Although we cannot exclude the possibility that mutation in the primary sequence or in the regulatory elements of the *mdr* gene are involved in resistance *in vivo*, our results strongly suggest that increased expression of the unaltered *mdr* gene is a likely mechanism by which tumour cells develop drug resistance. The nature and function of the mature polypeptide encoded by the cloned cDNA remains largely unknown. Recently, Riordan and co-workers have reported a cDNA (pcHP1) encoding part of the membrane glycoprotein gp170 (ref. 4). This glycoprotein is overexpressed in several multidrug-resistant hamster, mouse and human cell lines including LZ (refs 14, 15). The pcHP1 detects a 4.5-kb mRNA overexpressed in the resistant line C5. Hybridization studies indicate that the gene encoding this 4.5-kb mRNA is part of a multigene family amplified in these cells. Thus, *mdr* and *gp170* genes could be identical, but identification of a second class of cDNA clones in our library leaves open the possibility that *gp170* and *mdr* cDNA are closely related members of a small gene family.

The biochemical mechanism that underlies multidrug resistance has not been elucidated. It is generally believed that resistance is linked to decreased intracellular concentrations of the drugs. Data consistent with two alternatives have been presented: (1) a decreased rate of drug uptake¹⁶; (2) an increased rate of drug efflux¹⁷. The complete characterization of the

Fig. 4 Transcription analysis of cells transfected with pDREX4. Total cellular RNA (10 µg) from ADM-resistant clones 1A, 1, 3, 5, 6, 8, 9, 10 and from the control hamster line LR73 were electrophoresed in a 1% agarose gel (a) containing 7% formaldehyde, 0.02 M MOPS (pH 7.0), 0.05 M sodium acetate and 0.001 M EDTA. The RNA was transferred to a re-usable hybridization membrane (GeneScreen Plus, Dupont) and the blot was probed with 5 × 10⁶ c.p.m. of ³²P-labelled cDNA subclone pcDR1.3. Autoradiography was performed for 2 (a) and 48 h (b). The large arrow in a indicates the position of the major 6-kb transcript homologous with pcDR1.3. c, Total cellular RNA (5 µg) was denatured, serially diluted (1:1) and loaded onto a nitrocellulose membrane using a 96-well manifold (Hybridot, BRL) and the blot hybridized to 1 × 10⁶ c.p.m. of ³²P-labelled pcDR1.3. The RNA dilution factor (1 to 128) is indicated.



nucleotide sequence of the biologically active *mdr* cDNA should permit insight into the mechanistic basis of multidrug resistance and may help distinguish between the two models.

We thank N. Groulx for technical assistance; R. Kaufman (Genetics Institute) for plasmid p91023B; W. E. Mushynski and C. P. Stanners for critical review of the manuscript and for the cell line LR73 and the use of tissue-culture facilities; and J. D'Amico for secretarial assistance. Supported by grant MA 9626 from the Medical Research Council of Canada (MRC) to P.G. and by grants to D.E.H. from the National Institutes of Health and Ajinomoto Corporation. J.M.C. is supported by an NIH training grant (5T32HL07574). Y.B.N. was supported by a Fogarty International Postdoctoral Fellowship.

Received 6 May; accepted 1 August 1986.

- Gros, P., Croop, J., Roninson, I. B., Varshavsky, A. & Housman, D. E. *Proc. natn. Acad. Sci. U.S.A.* **83**, 337-341 (1986).
- Gros, P., Croop, J., Mukayama, T., Fallows, D. A. & Housman, D. E. *J. cell. Biochem. Abstr.* **10A**, 51 (1986).
- Roninson, I. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **83**, 4538 (1986).
- Riordan, J. R. *et al. Nature* **316**, 817-819 (1985).
- Roninson, I. B. *Nucleic Acids Res.* **11**, 5413-5431 (1983).
- Roninson, I. B., Abelson, H. T., Housman, D. E., Howell, N. & Varshavsky, A. *Nature* **309**, 626-628 (1984).
- Ling, V. & Thompson, L. H. *J. cell. Physiol.* **83**, 103-111 (1974).
- Howell, N., Belli, T. A., Zaczekiewicz, L. T. & Belli, J. A. *Cancer Res.* **44**, 4023-4030 (1984).
- Fojo, A. T., Whang-Peng, J., Gottesman, M. M. & Pastan, I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 7661-7664 (1985).
- Kaufman, R. J. & Sharp, P. A. *Molec. cell. Biol.* **2**, 1304-1319 (1982).
- Wong, G. G. *et al. Science* **228**, 810-814 (1985).
- Pollard, J. W. & Stanners, C. P. *J. cell. Physiol.* **98**, 571-586 (1979).
- Shen, D. W. *et al. Science* **232**, 643 (1986).
- Kartner, N., Riordan, J. R. & Ling, V. *Science* **221**, 1285-1289 (1983).
- Gerlach, J., Kartner, N., Bell, D. R. & Ling, V. *Cancer Surv.* (in the press).
- Ling, V. *et al. Cancer Treat. Rep.* **67**, 869-874 (1983).
- Inaba, M. *et al. Cancer Res.* **39**, 2200-2203 (1979).
- Norrander, J., Kempe, T. & Messing, J. *Gene* **26**, 101-106 (1983).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **132**, 61-13 (1983).
- Grosfeld, F. G. *et al. Nucleic Acids Res.* **10**, 6715-6731 (1982).
- Paige, C. J., Kincade, P. N. & Ralph, P. J. *Immun.* **121**, 641-648 (1978).
- Gubler, V. & Hoffman, B. J. *Gene* **25**, 263-269 (1983).
- Benton, W. D. & Davis, R. W. *Science* **196**, 180-182 (1977).
- Wigler, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 1373-1377 (1979).

A 3' enhancer is required for temporal and tissue-specific transcriptional activation of the chicken adult β -globin gene

Ok-Ryun Choi & James Douglas Engel*

Department of Biochemistry, Molecular Biology & Cell Biology,
2153 Sheridan Road, Northwestern University, Evanston,
Illinois 60201, USA

The chicken adult β -globin gene is one of the more intensively investigated developmentally regulated loci in higher eukaryotes. Detailed molecular analysis of the locus^{1,2} allows precise examination of the chromosomal changes that occur on activation of the gene during erythroid maturation. The best studied of these changes are the acquisition of DNase I hypersensitivity^{3,4}, developmentally correlated alteration of CpG-specific cytosine methylation patterns^{5,6} and *in vitro* assembly of erythroid-specific protein complexes 5' to the gene that mimics *in vivo* creation of the 5' DNase I hypersensitive 'region' lying 60 to 260 nucleotides 5' to the β -globin cap site in red blood cell chromatin⁷⁻⁹. Here we demonstrate that proximal β -globin DNA sequences lying >112 base pairs (bp) 5' to the cap site are not involved in determining the erythroid-specific induction characteristics of this gene in transient expression assays, whereas an enhancer sequence within

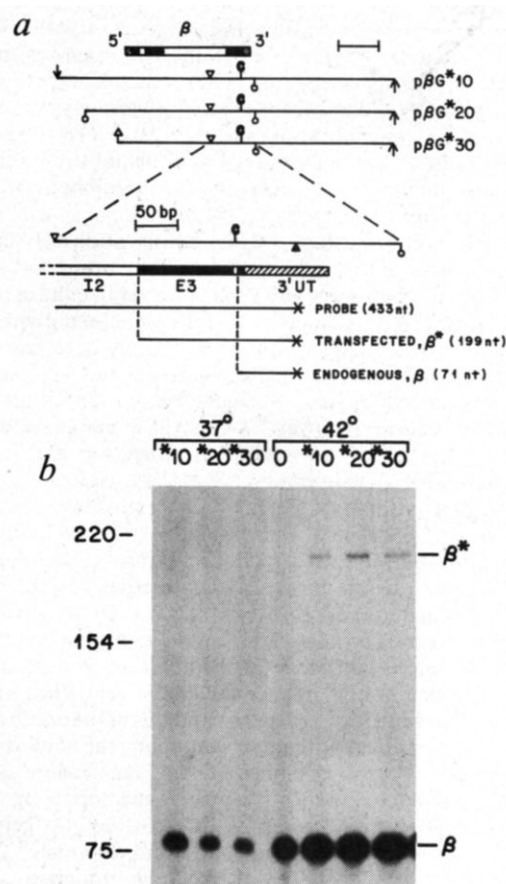


Fig. 1 5'-flanking sequence requirements for β -globin gene transcription. *a*, Globin gene plasmid constructions; *b*, S₁ nuclease protection of mRNA transcribed from endogenous and transfected β -globin genes. *a*, p β G*10 is identical to plasmid p β 1BR15, except that an 8 bp *Cla*I linker has been inserted at an *Rsa*I site within the β -globin gene (nucleotide +1,412)². p β G*20 is a 5' deletion of p β G*10 (containing only 511 nucleotides of 5'-flanking sequence) and p β G*30 is a 5' deletion in which all but 112 bp of DNA sequence next to the β -globin cap site has been deleted. Below, a diagram showing the 3' end of the structural gene, the S₁ nuclease protection probe and expected protection fragments from mRNA transcribed from either the episomal (β^*) or endogenous (β) globin genes. The S₁ probe used in these experiments was a restriction fragment of the β^* -globin gene (containing the *Cla*I linker) of 433 bp from the *Hind*III site (position +1058 in the β -globin second intron) to an end-labelled *Msp*I site at position +1,483 in the 3' untranslated region of the β -globin gene third exon². Protection of β -globin mRNA transcribed from the endogenous gene then yields a fragment of 71 nucleotides (β), while protection of a transcript from the transfected gene is expected to yield a 199 nucleotide band (β^*). RNA:DNA (probe) hybridizations, S₁ nuclease digestion and denaturing polyacrylamide gel electrophoreses were performed as previously described^{32,33}. Nt, nucleotides. Restriction enzyme recognition site symbols: *Eco*RI, $\downarrow$; *Hind*III, ∇ ; *Cla*I, $\uparrow$; *Sac*I, $\diamond$; *Bam*HI, $\uparrow$; *Msp*I, $\blacktriangle$. Bar, 500 bp. *b*, β^* -globin plasmids depicted above were transfected into HD3 cells and the recovered RNA was analysed by S₁ nuclease protection. (Lane 0, mock-transfected HD3 cell RNA protection products after propagation at 42°C.) DNA fragment sizes (left-hand column) in nucleotides.

a 300-bp *Pvu*II fragment lying ~400 nucleotides 3' to the polyadenylation signal is intimately involved in determining the erythroid cell specificity and correct time of induction of β -globin transcription during red cell maturation.

Modified β -globin gene-containing plasmids were tested for the ability to transcribe β -globin messenger RNA after transfection into both HD3 cells (an inducible avian erythroid precursor cell line) and chick embryo fibroblasts. HD3 cells are a tem-

* To whom correspondence should be addressed.

perature-sensitive avian erythroblastosis virus (ts34AEV) transformed line which, when shifted to the virally non-permissive temperature in the presence of erythropoietin, undergoes terminal erythroid differentiation to produce erythroid cells phenotypically resembling mature (adult) chicken red blood cells¹⁰. Weintraub and colleagues¹¹ have demonstrated that the β -globin gene in HD3 cells becomes transcriptionally activated only after thermal induction of differentiation.

HD3 cells were transfected using modifications of standard protocols using diethyl-aminoethyl (DEAE)-dextran¹². After transfection, the cells were allowed to recover in culture for 12 h at 35 °C in complex medium¹³. The cells were then divided into two aliquots: one was grown at 35 °C (virally permissive); the other was shifted to the restrictive temperature of 42 °C to allow globin induction after partial cellular maturation on the erythroid differentiation pathway¹⁰. After 36 h viable cells were counted, collected by low-speed centrifugation and lysed in guanidinium thiocyanate solution; RNA was isolated by pelleting through a cushion of CsCl^{14,15}. An equivalent amount of RNA was used for each S₁-nuclease protection, normalized for viable cell number. In each case, transcription of the endogenous β -globin gene was monitored as an internal control for the extent of induction (Fig. 1 legend). We could distinguish between transcripts synthesized from the endogenous (β) gene and the exogenous (β^*) gene by the addition of a small oligonucleotide linker (which 'marks' the globin gene third exon) in the transfected gene. The 5'-deletion mutants of the entire 4.2 kbp genomic *Bam*HI/*Eco*RI fragment containing the adult β -globin gene² (pBG*10) were examined in the transfection assay to determine whether flanking sequences encompassing the β -globin 'hypersensitive domain' 5' to the structural gene⁴ can confer β^* -globin erythroid inducibility, and where such sequences might be localized. Two 5'-flanking sequence deletion mutants from pBG*10, which contains all sequence information both 5' and 3' to the β -globin gene were made. These have either 511 bp of 5'-flanking sequence (pBG*20) or 112 bp of 5'-flanking sequence (pBG*30; Fig. 1a). The deletion mutants were transfected into HD3 cells and induced to allow erythroid differentiation. RNA from these transfected cells was then analysed for the presence of RNA transcribed from the exogenous β^* -globin gene by S₁ nuclease protection.

Surprisingly, the genes in all three plasmid constructions (pBG*10, pBG*20 and pBG*30) are transiently expressed at comparable efficiency, and all three transfected genes are erythroid-inducible (Fig. 1b). This implies that the response of β -globin messenger RNA synthesis during erythroid cell maturation is principally mediated by DNA sequences lying in and/or 3' to the structural gene locus, and that only minimal 5'-flanking sequence (<112 bp 5' to the β -globin cap site) need be present to allow erythroid-inducible expression of transfected DNA. This result is therefore consistent with observations analysing the expression of the mouse β -globin gene in transformed MEL cells^{16,17}.

Recombinant pBG*30 clearly lacks most of the adult β -globin DNase I 'hypersensitive domain' lying 60–260 bp 5' to the gene. The structural assembly of this domain correlates very accurately with the inducible expression of the chicken adult β -globin gene in erythroid cells^{8,9}. What is the role of this domain in the light of our results? First, it may be that this region is not involved in induction. We think this is unlikely for two reasons: (1) the time of appearance of the hypersensitive structure correlates well with the onset of β -globin transcription¹¹ and (2) the structure can only be reconstituted *in vitro* on cloned DNA using nuclear extracts from cells which express the β -globin gene⁹. We believe that the state of the DNA being examined is more likely to provide the explanation. Perhaps the endogenous chromosomal gene is structurally constrained in a way that the episomal gene is not; or, equally likely, perhaps the hypersensitive structure is associated with a regulatory function of the gene which is irrelevant to the time course and the tissue-specificity.

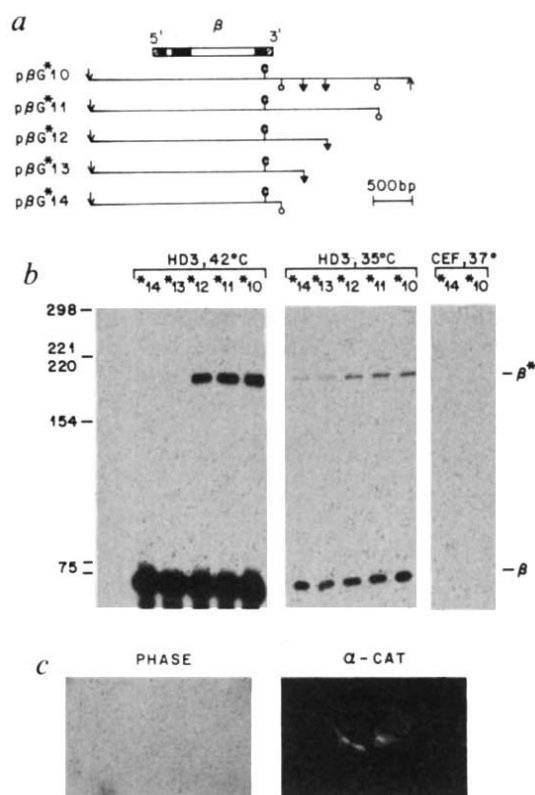


Fig. 2 3' Flanking sequence requirements for β -globin gene transcription. **a**, 3' Deletion constructions; **b**, S₁ nuclease protection analysis of 3' deletion plasmids; **c**, immunofluorescence of transfected chick embryo fibroblasts. **a**, Plasmid pBG*10 was deleted at restriction sites 3' to the β^* -globin gene to form the plasmids shown (pBG*11–14), which vary only in the distance of the β -globin mRNA from the polyadenylation site¹⁸: pBG*14, +64 bp; pBG*13, +354 bp; pBG*12, +650 bp; pBG*11, +1,330 bp; pBG*10, 1,770 bp. Restriction enzyme recognition site symbols as in Fig. 1, legend. HD3 cells were transfected, split 1:1 and propagated at permissive and non-permissive temperatures as described; RNA was isolated, hybridized to the S₁ probe shown in Fig. 1a, and treated with S₁ nuclease as before. The arrows depict protection of mRNA transcribed from the transfected (β^*) and endogenous (β) genes. In the same experiment, primary chicken embryo fibroblasts were separately cotransfected with β -globin plasmids pBG*10 and pRSVCat or pBG*14 and pRSVCat^{19,22}. After 48 h RNA was isolated from the transfected fibroblasts and analysed for the presence or absence of β^* -globin mRNA as above. Left-hand column, DNA fragment size. **c**, As a positive control for transfection efficiency, part of the culture of transfected CEF used for RNA isolation in **b** was examined by direct fixation and antibody staining as an assay for chloramphenicol acetyltransferase (CAT) expression. Left, phase photomicrograph of the transfected CEF culture; right, the same field in epifluorescence after reaction with rabbit anti-CAT antiserum and goat anti-rabbit FITC-labelled IgG. Labelling efficiency in this experiment was ~10%.

We attempted to localize the *cis*-acting regulatory sequence effect. As sequences controlling erythroid inducibility were not found at the 5' end of the gene, we next examined constructions containing serial deletions of DNA sequence information from 3' flanking sequences distal to the β^* -globin polyadenylation site (pBG*11–14; Fig. 2a)^{2,18}. Figure 2b shows that at the virally permissive temperature (35 °C) low levels of transcription are detected in each transfection construct (as revealed by the presence of a weak S₁ protection band of 199 bases for β^* transcripts). The endogenous β -globin gene in these cells is also transcribed only at low levels. After induction of erythroid maturation by shift to 42 °C for 36 h, only cells transfected with pBG*10, 11 or 12 inducibly transcribe, splice and accumulate the newly introduced β^* -globin gene, whereas those transfected

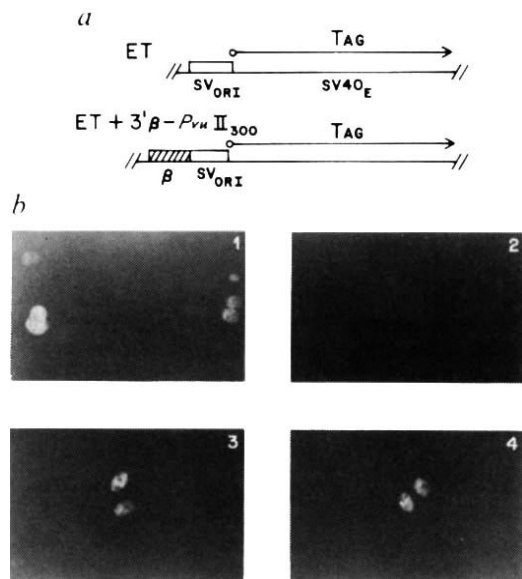


Fig. 3 The β -globin enhancer lies in a 300-bp *PvuII* fragment 3' to the β -globin gene and confers inducible erythroid specificity. *a*, SV40 T-antigen/ β -globin enhancer test constructions; *b*, nuclear fluorescent staining; the 300 bp β -globin 3' *PvuII* fragment enhances T-antigen expression—only in induced HD3 cells. *a*, A plasmid containing intact SV40 origin and early transcription regulatory sequences and the coding sequences for T antigen (TAG), but without the SV40 72-bp repeats (the 'enhancer trap'; ET²⁰) was ligated to the 300-bp *PvuII* fragment (β) required for β^* -globin mRNA accumulation as defined by the experiments shown in Fig. 2*b*. This new plasmid (ET+3' β -*PvuII*₃₀₀) and the ET parental plasmid were separately transfected into HD3 cells or CEF and examined for nuclear T-antigen accumulation (below). *b*, HD3 cells were transfected with both plasmids shown above, split into equal aliquots and incubated at either 35 or 42°C. No fluorescence was detected in cells transfected with ET plasmid alone at either temperature (data not shown), whereas T-antigen was abundantly expressed as assayed by nuclear fluorescent staining in 42°C induced cells (1), but not in 35°C uninduced cells (2; see text). Low-level nuclear T-antigen fluorescence is detectable in CEF after transfection with either ET+3' β -*PvuII*₃₀₀ (3) or ET alone (4). Panels 1–4, fields in epifluorescence after staining with anti-T-antigen monoclonal antibody and FITC-labelled goat anti-mouse IgG antibody. Exposure times: panels 1 and 2, 45 s; panels 3 and 4, 3 min.

with pBG*13 and 14 (deleting most of the β^* -globin 3'-flanking sequence) do not (Fig. 2*b*). Thus both the transfected gene in pBG*12 and the endogenous gene are induced on shifting the HD3 cells into the erythroid differentiation pathway, implying that 3' DNA sequences, present in pBG*12 but not in pBG*13, are required for stable accumulation of β^* -globin mRNA from plasmid pBG*12. To determine whether this 300-bp *PvuII* fragment is required for maintenance of transcription or is involved in determining the erythroid specificity during induction, constructions including (pBG*10) and lacking (pBG*14) the 300-bp *PvuII* fragment were transfected into chick embryo fibroblasts (CEF). CEF transfected with either construct express β^* -globin mRNA only at very low levels (Fig. 2*b*), showing that although plasmid pBG*10 contains sufficient sequence information for erythroid-specific induction (Figs 1*b* and 2*b*), neither globin gene construct is efficiently expressed in non-erythroid cells. We estimate from a comparison of β^* -globin signal intensity (Fig. 2*b*; CEF lanes) that the difference in accumulated levels of β^* -globin mRNAs is 20- to 50-fold when comparing induced erythroid cells with CEF.

To ensure that the differences in accumulated mRNA levels were not caused by different transfection efficiencies of erythroid and non-erythroid cells, we cotransfected the β^* -globin plasmids and pRSVCat¹⁹ into CEF cells (Fig. 2*c*). After transfection

```

10      20      30      40      50      60
CTGGGTGGGG GCAGGTTGCA GATAACATT TTGCTATCAA GACTTGCACA GACCTTGTGT
EcoP1

70      80      90      100     110     120
ATGCACCTTC TCACCCCTACG CTGCCCTTC TGCTGCTCTG CGTGAGGGA GAGAGGGGGT
MboII      HphI      EcoP15      MnlI      MboII      MnlI

130     140     150     160     170     180
TAATCCTGTC AATAGTAGTA ACTATTAATG ATGTTCTGGA ACTAAAAATC AATTGGTGTG
EcoRI*

190     200     210     220     230     240
ATTTCATGTC AAACATATAGG GGTGTTTAAA TGTGCACAGC GCAGGGCTCC ATAGGGTGGG
SphI      HgiAI      HhaI

250     260     270     280     290     300
GGTGGTCTTT GGCATGGGAN AACCACTCAT CCCATTGAGA TGATCAGAGG GCAGCAAAAG
BclI      Sau3AI

.303
CAG

```

Fig. 4 Sequence of the 3' β -globin enhancer. The *PvuII* fragment shown to confer erythroid tissue-specific enhancement (Figs 2, 3) was subcloned into plasmid pGEM3 (Promega Biotec, Madison), linearized by digestion with *ScaI* restriction nuclease and directly sequenced from the heat-denatured plasmid after hybridization with T7 and Sp6 promoter-specific primers followed by normal dideoxy-sequencing methods. The sequence is depicted in the same orientation as the mRNA-sense DNA strand.

the cells were examined by indirect immune fluorescence for the CAT protein using a rabbit anti-CAT antiserum. About 10% of the cells react with the anti-CAT antibody in a CaPO₄ cotransfection experiment of pBG*10 with pRSVCat (Fig. 2*c*). CEF from the same transfection were used for both the immunofluorescence experiment (Fig. 2*c*) and the S₁ protection analysis (Fig. 2*b*). Thus, although the β^* -globin gene has been transfected into the CEF cells, these cells are unable efficiently to use a β^* -globin DNA template for transcription.

To examine the requirement for the 300-bp *PvuII* fragment 3' to the β -globin gene for mRNA accumulation further, we transfected HD3 cells with the simian virus 40 (SV40) 'enhancer trap'²⁰ with and without the 300-bp *PvuII* fragment linked in *cis* (Fig. 3*a*). The enhancer trap (ET), an SV40 vector lacking the viral 72-bp SV40 enhancer repeats, accumulates T-antigen only when joined to an exogenous enhancer and transfected into cells of that developmental lineage in which the enhancer is active^{20,21}. After transfection, half the cells were grown at the virally permissive temperature (35°C), while the other half were grown at the non-permissive temperature (42°C) to allow β -globin mRNA accumulation. After 36 h, the cells were stained with an anti-T-antigen monoclonal antibody followed by fluorescein-labelled (FITC) goat anti-mouse IgG. Fluorescence was quite intense in ~5% of cells transfected with the ET plasmid containing the 3' β -globin DNA fragment and grown at 42°C (Fig. 3*b*, panel 1); only background fluorescence was seen in most cells in the culture that was not differentiation-induced (Fig. 3*b*, panel 2) as might be anticipated from the results shown in Figs 1 and 2. No fluorescence was detected in HD3 cells (at either temperature) when transfected with the unmodified ET plasmid (data not shown). The results of this experiment are consistent with similar observations by F. Weber and W. Schaffner (personal communication).

To ascertain whether the β -globin 3' *PvuII* fragment could also enhance T-antigen expression in non-erythroid cells, the plasmids shown in Fig. 3*a* were separately transfected into CEF²² and examined 48 h later for accumulation of T-antigen in the nuclei. Low-level fluorescent nuclear staining was detectable with both plasmids (Fig. 3*b*, 3, 4) but no difference in intensity was seen between the fluorescence of cells transfected with plasmid containing and lacking the *PvuII* fragment, and the fluorescence was far less intense than in the differentiation-induced HD3 cells (Fig. 3*b*, panel 1). Thus an erythroid-specific enhancer exists within the 300-bp 3' *PvuII* DNA fragment adjacent to the gene, which is activated during chick erythroid cell maturation with the same time-course as β -globin mRNA induction. Furthermore, the enhancer displays these effects independently of other β -globin locus DNA sequences.

The DNA sequence of the enhancer activity localized in the 303-bp 3' β -globin *PvuII* fragment (Fig. 4) is only slightly similar to the 'core' enhancer consensus sequence (GTGGTTTGG)²³, or to other cellular enhancers. A precise 'core' sequence match can be found centred 335 nucleotides 5' to the β -globin capping site² but is functionally inert in our assays. There are some reasonably strong, although short, sequence homologies to the *PvuII* fragment in current databases (for example, to a murine κ V-region germline flanking repetitive DNA sequence²⁴), but no functional similarity is known.

Although no other cellular gene is yet known to be associated with a 3' tissue-specific enhancer element, the location of the enhancer identified here, when taken together with analysis of the location of immunoglobulin (Ig) enhancers inside their structural genes^{21,25} implies that the characterization of genes only by examination of 5' flanking sequences may miss some important regulatory regions. Furthermore, the fact that the β -globin enhancer identified here exerts its regulatory effect only after commitment to terminal erythroid differentiation suggests that similar developmental specificity may be found in other differentially controlled genes. As an example, this hypothesis predicts that enhancer activation in the Ig gene loci may not be a direct result of rearrangement *per se*, but of progressive maturation from a cell type in which the Ig heavy-chain enhancer is inactive (in the predecessor of the immature pre-B or null cell, containing unrearranged Ig loci) to the developmentally more mature cell types (null and pre-B) in which some heavy-chain gene segments have already recombined²⁶. Recently, Gerster *et al.*²⁷ have found roughly equal transcription rates within Ig heavy-chain genes in transformed pre-B (partially rearranged) and myeloma (productively rearranged) lymphocyte lines.

Why is the β -globin enhancer located 3' to the gene that it activates? There are several possible mechanisms for the activity of such a uniquely localized element. First, the chicken β -globins are unique in their organization among known vertebrate β -globin gene clusters, in that the ϵ gene lies at the extreme 3' end of the gene cluster¹. We have presented evidence that this gene arose from an ancestral adult β -globin by an evolutionarily recent gene conversion event²⁸. Perhaps the enhancer defined here was situated between these two ancestral adult β -globins (now the adult β and embryonic ϵ genes) to potentiate expression from both promoters, as the proximity of enhancers to promoters does display an effect on the level of accumulated gene products²⁹. Another possibility is that the enhancer suppresses ('silences')³⁰ the ϵ -globin gene 3' to it at the same time as enhancing the β -globin gene.

After our manuscript was submitted, Hesse *et al.*³¹ reported observations showing that sequences 3' to the β -globin gene are required for expression of β -globin promoter/*cat* gene chimaeras after transfection into primary chicken erythrocytes. The 3' sequence required for *cat* activity overlaps the DNA sequence shown in Fig. 4.

We thank C. Trainor and J. Gallarda for discussions, Y. D. Choi for help in the immunofluorescence experiments, K. Rundell (N.U. Med. School) and D. Standring (U.C.S.F.) for gifts of antibodies, and W. Schaffner for communication of results before publication and the enhancer trap vector. Supported by NIH and the American Cancer Society (grant numbers HL24415, GM28896 and NP-528), the NIH Fogarty Center (TW01016), and the American Cancer Society (Faculty Research Award).

Received 3 March; accepted 8 August 1986.

1. Dolan, M., Sugarman, J., Dodgson, J. B. & Engel, J. D. *Cell* **24**, 669-677 (1981).
2. Dolan, M., Dodgson, J. B. & Engel, J. D. *J. biol. Chem.* **258**, 3983-3990 (1983).
3. Stalder, J. *et al. Cell* **20**, 451-460 (1980).
4. McGhee, J. D., Wood, W. I., Dolan, M., Engel, J. D. & Felsenfeld, G. *Cell* **27**, 45-55 (1981).
5. McGhee, J. D. & Ginder, G. D. *Nature* **280**, 419-420 (1979).
6. Ginder, G. D. & McGhee, J. D. in *Organization and Expression of Globin Genes* (eds Stamatoyannopoulos, G. & Nienhuis, A.) 191-201 (Liss, New York, 1981).
7. Nickol, J. M. & Felsenfeld, G. M. *Cell* **35**, 467-477 (1983).

8. Emerson, B. M. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **81**, 95-99 (1984).
9. Emerson, B. M., Lewif, C. D. & Felsenfeld, G. *Cell* **41**, 21-30 (1985).
10. Beug, H., Doederlein, G., Freudenstein, C. & Graf, T. *J. cell. Physiol. suppl.* **1**, 195-207 (1982).
11. Weintraub, H., Beug, H., Groudine, M. & Graf, T. *Cell* **28**, 931-940 (1982).
12. de Villiers, J. & Schaffner, W. in *Nucleic Acid Biochemistry Techniques in the Life Sciences* (ed. Flavell, R. A.) 1-20 (Elsevier, Amsterdam, 1983).
13. Beug, H. *et al. Virology* **145**, 141-153 (1985).
14. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
15. Glisin, V., Crkvenjakov, R. & Byus, C. *Biochemistry* **13**, 2633-2637 (1974).
16. Charnay, P. *et al. Cell* **38**, 251-263 (1984).
17. Wright, S., Rosenthal, A., Flavell, R. & Grosfeld, F. *Cell* **38**, 265-273 (1984).
18. Richards, R. I., Shine, J., Ulvrich, A., Wells, J. R. E. & Goodman, H. M. *Nucleic Acids Res.* **7**, 1137-1146 (1979).
19. Gorman, C. M., Merlino, G. T., Willingham, M. C., Pastan, I. & Howard, B. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 6777-6781 (1982).
20. Weber, F., de Villiers, J. & Schaffner, W. *Cell* **36**, 983-992 (1984).
21. Banerji, J., Olsen, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
22. Choi, O.-R., Trainor, C., Graf, T., Beug, H. & Engel, J. D. *Molec. cell. Biol.* **6**, 1751-1759 (1986).
23. Weiher, H., Konig, M. & Gruss, P. *Science* **219**, 626-631 (1983).
24. Gebhard, W., Meitinger, T., Hochtl, J. & Zachau, H. G. *J. molec. Biol.* **157**, 453-471 (1982).
25. Gillies, S., Morrison, S. L., Oi, V. T. & Tonegawa, F. *Cell* **33**, 717-728 (1983).
26. Yancopoulos, G. D. & Alt, F. W. *Cell* **40**, 271-281 (1985).
27. Gerster, T., Picard, D. & Schaffner, W. *Cell* **45**, 45-52 (1986).
28. Dodgson, J. B. *et al. J. biol. Chem.* **258**, 12685-12692 (1983).
29. Kadesch, T. & Berg, P. *Molec. cell Biol.* **6**, 2593-2601 (1986).
30. Brand, A. H., Breiden, L., Abraham, J., Sternglanz, R. & Nasmyth, K. *Cell* **41**, 41-48 (1985).
31. Hesse, J. E., Nickol, J. M., Lieber, M. R. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **83**, 4312-4316 (1986).
32. Weaver, R. F. & Weissman, C. *Nucleic Acids Res.* **7**, 1175-1193 (1979).
33. Fischer, H. D., Dodgson, J. B., Hughes, F. H. & Engel, J. D. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2733-2737 (1984).

Complete protein sequence and identification of structural domains of human apolipoprotein B

T. J. Knott*, R. J. Pease*, L. M. Powell*, S. C. Wallis*, S. C. Rall Jr†, T. L. Innerarity†, B. Blackhart†, W. H. Taylor‡, Y. Marcel§, R. Milne§, D. Johnson||, M. Fuller||, A. J. Lusis||, B. J. McCarthy†, R. W. Mahley†, B. Levy-Wilson† & J. Scott*¶

* Molecular Medicine Research Group, MRC Clinical Research Centre, Watford Road, Harrow HA1 3UJ, UK

† Gladstone Foundation Laboratories, Cardiovascular Research Institute, Departments of Medicine, Pathology and Pharmaceutical Chemistry, University of California, San Francisco, California 94140, USA

‡ Department of Crystallography, Birkbeck College, University of London, London WC1E 7HX, UK

§ Clinical Research Institute of Montreal, 110 West Pine Avenue, Montreal, Quebec H2W 1R7, Canada

|| Department of Medicine, University of California, Los Angeles, California 90024, USA

Epidemiological, pathological and genetic studies show a strong positive correlation between elevated plasma concentrations of low-density lipoprotein (LDL) cholesterol and the risk of premature coronary heart disease¹⁻³. Apolipoprotein (apo) B-100 is the sole protein component of LDL and is the ligand responsible for the receptor-mediated uptake and clearance of LDL from the circulation²⁻⁶. Apo B-100 is made by the liver and is essential for the assembly of triglyceride-rich very low-density lipoproteins (VLDL) in the cisternae of the endoplasmic reticulum and for their secretion into the plasma. VLDL transports triglyceride to peripheral muscle and adipose tissue, where the triglyceride is hydrolysed by lipoprotein lipase. The resultant particle, relatively enriched in cholesteryl ester, constitutes LDL. LDL delivers cholesterol to peripheral tissues where it is used for membrane and steroid hormone biosynthesis and to the liver, the only organ which can catabolize and excrete cholesterol. Plasma LDL levels are therefore determined by the balance between their rate of production from VLDL and clearance by the hepatic LDL (apo B/E) receptor pathway. Here we report the complete 4,563-amino-acid sequence of apo B-100 precursor (relative molecular mass (M_r) 514,000 (514K)) determined from complementary DNA

¶ To whom correspondence should be addressed.

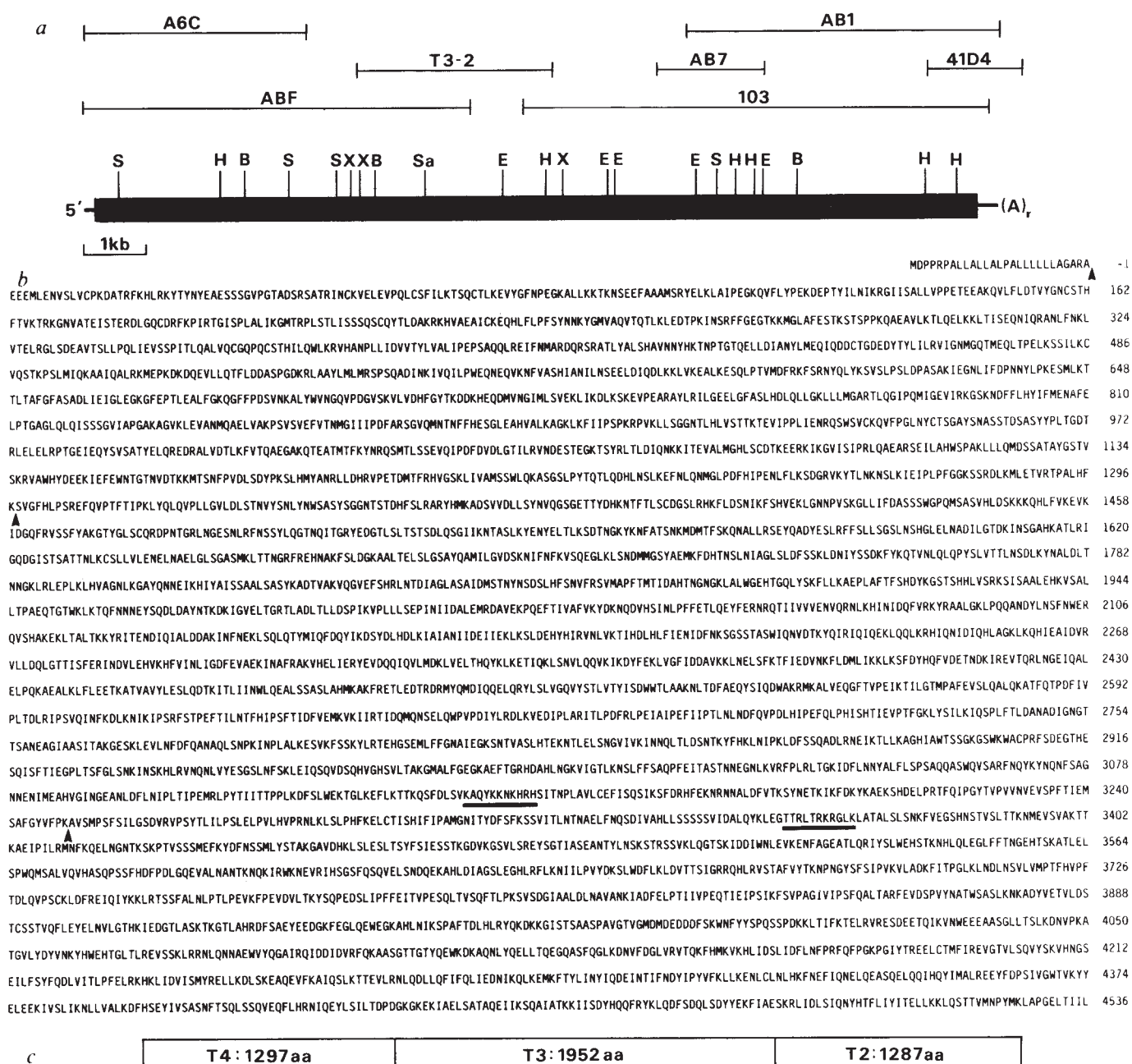


Fig. 1 **a**, Restriction map of apo B-100 cDNA. Abbreviations, Sa, *Sal*I; H, *Hind*III; B, *Bam*HI; X, *Xba*I; S, *Sst*I; E, *Eco*R1. **b**, Amino-acid sequence of human apolipoprotein B-100. Vertical arrows, mature protein amino terminus (residue 1) and two thrombin cleavage sites. Solid bars, two stretches of basic residues (see text). The four proline-rich clusters occur at amino acids 1,167–1,324, 2,561–2,726, 3,102–3,305 and 3,687–3,865 (see text). **c**, Thrombolytic cleavage fragments of apo B-100. **a**, Apo B-100 mRNA comprises 14,121 nucleotides and is therefore the longest single mRNA characterized to date⁷. The 5' untranslated region is 128 nucleotides. There is a single long open reading frame encoding 4,563 amino acids. The 3' non-coding region is 304 bp in length.

Methods. **a**, Isolation and DNA sequencing of clones AB1, AB7 and 41D4 encoding the carboxyl-terminal thrombolytic peptide (T2) and apo B-100 has been described³⁰. We isolated the cDNAs spanning T4 and T3 as follows. A *λ*gt11 expression library prepared from HepG2 poly(A)⁺RNA was screened with polyclonal antibodies against human LDL. One positive clone (A6C) containing a 3.5-kilobase (kb) insert also hybridized to synthetic oligonucleotides complementary to the amino-terminal sequence of the mature protein^{30,37} and the apo B-100 peptide R3-1 (ref. 38). The longer clone, ABF, was subsequently obtained from a size-selected adult liver cDNA library³⁹. The overlapping clones T3-2 and 103 were isolated from a HepG2 cDNA library (made by oligo-dT priming on poly(A)⁺RNA and cloned in *λ*gt10) using existing cDNA fragments as probes. **b**, The amino terminus of the mature plasma protein Glu-Glu-Glu-Glu has been determined by direct sequencing^{30,37}. We have previously shown that a carboxyl-terminal peptide isolated from a tryptic digest of intact LDL and sequenced by mass spectrometry corresponds to the last 10 residues of the predicted apo B-100 precursor, establishing that there is no processing at the carboxyl-terminus³⁰. The mature apo B-100 polypeptide therefore comprises 4,536 amino acids. Partial amino-acid sequences have been obtained from 32 peptides derived by thrombolytic, tryptic, V8 protease or CNBr-cleavage of apo B-100 and represent sequences beginning at residues 1, 5, 100, 146, 205, 271, 846, 875, 1,298, 1,439, 2,016, 2,138, 2,144, 2,496, 2,981, 3,085, 3,109, 3,127, 3,134, 3,250, 3,308, 3,356, 3,564, 3,570, 3,607, 3,640, 3,659, 3,720, 3,820, 3,842, 4,235 and 4,356, confirming the protein structure reported here. **c**, Apo B-100 in LDL is specifically cleaved *in vitro* at two sites by both thrombin and kallikrein⁸. Sequence analysis of apo B-100 thrombolytic peptides shows that the sites are Lys-1,297 and Lys-3,249. Residues 1–1,297 correspond to peptide T4 of Cardin *et al.*⁹, 1,298–3,249 to peptide T3, and 3,250–4,536 to peptide T2. Peptide T4 is indistinguishable from apo B-26, a naturally occurring proteolytic degradation product of apo B-100 in LDL⁴⁰.

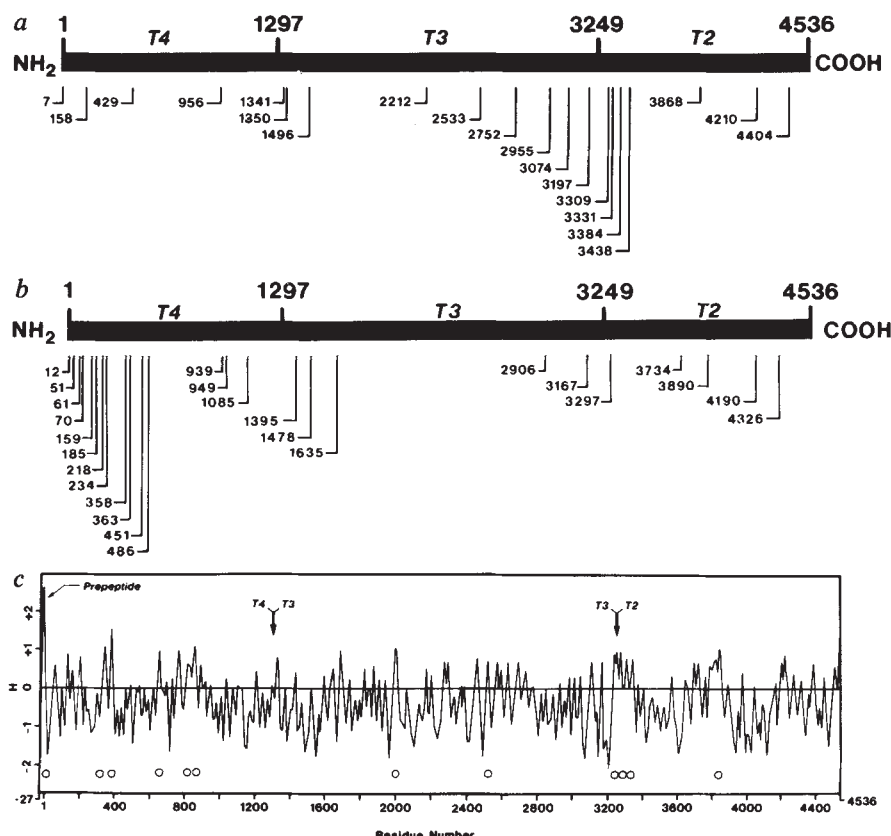


Fig. 2 *a*, Potential *N*-glycosylation sites in apo B-100. Five high mannose chains and an equal or slightly greater number of the complex type are found in each apo B-100 molecule²³, therefore not all 20 potential sites are used. Comparison of the predicted *M_s* of apo B-100 thrombolytic fragments (see Fig. 1c) with their migration rates in SDS-polyacrylamide gels (T4, 145K versus 145K; T3, 221K versus 238K; T2, 146K versus 170K) suggests that T3 and T2 are glycosylated but that T4 is not (data not shown). Direct protein sequencing has shown that the *N*-glycosylation site at residue 7 of apo B-100 is not used but the first two sites in T2 (amino acids 3,309 and 3,331) are glycosylated, since no signal could be obtained at these positions³⁰. *b*, Position of cysteine residues in apo B-100. *c*, Hydrophobicity profile of apo B-100. The hydrophobicity index (H) of Kyte and Doolittle¹⁵ was plotted against the apo B-100 amino-acid sequence using a windowing average of 18 residues. Further analysis by Eisenberg's hydrophobic moment plot⁴¹ identified the signal peptide as the only potential membrane spanning domain. Eleven other sequences (indicated as circles) are predicted to lie at the lipid/aqueous interface and are in the most strongly hydrophobic parts of apo B-100.

clones. Numerous lipid-binding structures are distributed throughout the extraordinary length of apo B-100 and must underlie its special functions as a nucleus for lipoprotein assembly and maintenance of plasma lipoprotein integrity. A domain enriched in basic amino-acid residues has been identified as important for the cellular uptake of cholesterol by the LDL receptor pathway.

Overlapping cDNAs (Fig. 1a) were isolated from liver and HepG2 cDNA libraries and their DNA sequence was determined (full sequence available on request from J.S.). Human apo B-100 is made as a 4,563-amino-acid precursor (Fig. 1b) from a messenger RNA of 14,121 nucleotides⁷. A hydrophobic 27-residue signal peptide is cleaved before secretion of nascent VLDL.

The mature protein of 4,536 amino acids has a predicted *M_r* of 512K compared with the 550K found by SDS-gel electrophoresis⁸. This discrepancy is probably caused by glycosylation as 8–10% of the mass of apo B-100 is *N*-linked oligosaccharide⁹. Twenty potential *N*-glycosylation sites (Asn-X-Thr/Ser) are distributed throughout the sequence with a cluster between residues 3,197–3,438 (Fig. 2a and legend).

The structure of human apo B has been analysed in terms of its functions in lipid binding, lipoprotein assembly and as the ligand responsible for LDL clearance by the LDL receptor pathway. Secondary structure predictions¹⁰ show that apo B has extensive α and β structure in close agreement with values derived from circular dichroism spectra¹¹. The profile is different from that of a 'typical' apolipoprotein which has a high α -helical content and almost no β -sheet. In most apolipoproteins lipid binding occurs through amphipathic α -helical segments^{12,13}. In apo B-100 few of the predicted α -helices are truly amphipathic in terms of charge distribution on the polar surface¹² except for one extended region (residues 2,000–2,600) which contains good examples of amphipathic α -helices, and may contribute to lipid binding.

Another structure capable of lipid interactions is the amphipathic β -strand that contains alternating hydrophobic and hydrophilic amino acids. Synthetic amphipathic β -strand peptides can organize lipids into LDL-like particles and can bind

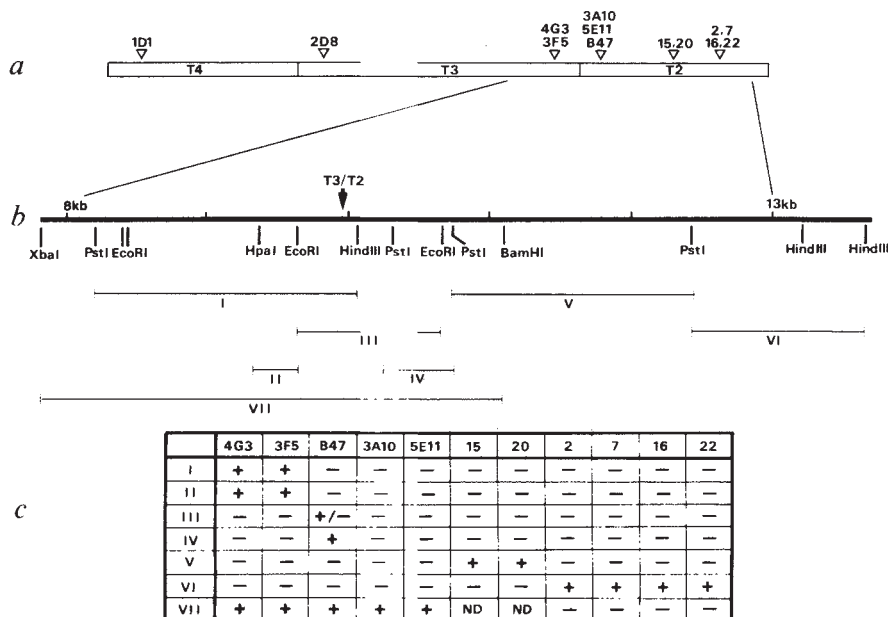
tightly to plasma LDL¹⁴. These structures occur throughout the apo B sequence, but particularly in four proline-rich clusters (Fig. 1b legend), where 42% of all proline residues are concentrated. These regions contain short repeated 5-amino-acid homologous sequences with the consensus acidic-aromatic-polar-aliphatic-proline. The repeats have high amphipathic β -sheet potential and are punctuated by proline residues which may participate in amphiphilic β -turn formation. These domains must have a significant impact on apo B-100 structure.

The Kyte-Doolittle hydrophobicity profile¹⁵ of apo B-100 resembles that of an integral membrane protein in that it has many hydrophobic regions some of which are extensive (Fig. 2c). Lipid-binding structures in the form of amphipathic α -helices, amphipathic β -strands and hydrophobic domains are distributed throughout the length of apo B-100 giving the molecule its characteristics (unique among the apolipoproteins) of insolubility in aqueous media and non-exchangeability¹⁶. Thus apo B-100 remains tightly associated with its core lipid throughout the metabolic transitions that lead to the formation of LDL, during which triglyceride and phospholipids are distributed to muscle and adipose tissue and all other VLDL apolipoproteins (apo E and the C apolipoproteins) are lost to different lipoprotein fractions.

Several features have been identified that may have importance for lipoprotein assembly. Cysteine residues are important in determining protein tertiary structure. Only 25 cysteines are present but 15 occur in the amino-terminal thrombolytic peptide T4 (Fig. 1c), 12 of those in the first 500 amino acids (Fig. 2b). In contrast, in T3 there is a sequence of 1,300 amino acids containing no cysteine. Six out of every seven cysteines in apo B-100 are involved in intramolecular disulphide linkage¹¹. Therefore at least one but probably not more than three occur as native cysteine (free sulphhydryl). Analysis of apo B-100 disulphide linkages indicates that T4 is not crosslinked to T3 or T2 (C. H. Spilman *et al.*, unpublished observations).

These results suggest that T4 is highly crosslinked within itself and possibly more globular in structure than the rest of the molecule. Potential *N*-glycosylation sites are apparently not

Fig. 3 Monoclonal antibody binding to apo B-100 fusion proteins. **a**, Epitope map for apo B-100. Isolation and characterization of monoclonal antibodies has been described elsewhere^{42,43}. Antibodies 1D1, 2D8, 4G3, 3F5, 3A10, 5E11 and B47 react against native LDL. Antibodies 2, 7, 15, 16, 20 and 22 were raised against fragments of apo B-100 and failed to recognize intact LDL. Monoclonals 3F5, 4G3, 3A10, 5E11 and B47 block LDL binding to the LDL receptor. The positions of epitopes for antibodies 1D1 and 2D8 have been determined elsewhere⁴². **b**, Fragments of apo B-100 cDNA were subcloned into pUR⁴⁴ or pING⁴⁵ expression vectors and used to transform *Escherichia coli*. Total bacterial protein from seven different constructs was isolated, subject to SDS-PAGE, transferred electrophoretically to nitrocellulose filters⁴⁶ then probed with anti-apo B antibodies and I¹²⁵. Protein A. **c**, Tabulation showing antibody binding to fusion proteins. Receptor blocking monoclonal antibodies 4G3 and 3F5⁴² react only with fusion proteins I (β -galactosidase + apo B residues 2,657–3,286) and II (ara B + apo B residues 3,029–3,132). Both epitopes therefore lie between residues 3,029 and 3,132. Antibody B47⁴³, also receptor blocking, reacts weakly with fusion protein III (β -galactosidase + apo B residues 3,133–3,494) but strongly with IV (β -galactosidase + apo B residues 3,350–3,506) suggesting that the epitope lies near the carboxy-terminus of fragment IV. The receptor blocking antibodies 3A10 and 5E11 recognize thrombolytic peptide T2 and compete with antibody B47 for binding to LDL (R.M. and Y.M., unpublished data). 3A10 and 5E11 react with fusion protein VII containing apo B-100 residues 2,488–3,636. Because proteins IV and V were negative with these antibodies, the epitopes probably include or span the boundary region of constructs IV and V (residue 3,506). Antibodies 15 and 20 react with fusion protein V (residues 3,506–4,082) whereas 2, 7, 16 and 22 react with fusion protein VI (residues 4,082–4,525).



used (Fig. 2a, legend), presumably because the complex tertiary structure of this part of the polypeptide makes them inaccessible. Many secretory and cell-surface proteins have similar cysteine-rich domains^{17,18} in which the disulphide bonds are thought to be formed by a disulphide isomerase at the time of translation^{19,20}. The ligand-binding domain of the LDL receptor is cysteine-rich and it has recently been demonstrated that deletions from this region prevent transport of the receptor from the rough endoplasmic reticulum to the Golgi apparatus²¹. By analogy the cysteine-rich domain of apo B-100 may confer a specific globular structure that is necessary for nascent lipoprotein assembly and transport from the endoplasmic reticulum to the Golgi apparatus.

Many lipid-binding structures are distributed at irregular intervals throughout apo B-100. Except for the amino terminus, the nascent protein may be a flexible backbone, tightly associated with the inner leaflet of the endoplasmic reticulum, forming a nucleus to which triglyceride and phospholipid are added from their sites of synthesis (the smooth endoplasmic reticulum) during lipoprotein assembly. Support for this comes from ultrastructural studies: a pool of apo B can be immunologically demonstrated in intestinal rough endoplasmic reticulum in the fasting state, but lipoprotein assembly is observed only in the non-fasting state and in the smooth endoplasmic reticulum²². Before secretion the nascent VLDL particle must undergo modification of high mannose carbohydrate side chains as apo B-100 is known to contain complex carbohydrate groups which are synthesized in the Golgi apparatus²³. However, inhibition of glycosylation with tunicamycin indicates that in the chick glycosylation is not essential for the synthesis or secretion of VLDL²⁴.

The receptor binding domain of apo B-100 is not available in VLDL, but becomes exposed after hydrolysis of VLDL triglyceride and the loss of other VLDL apoproteins to different lipoprotein fractions. VLDL hydrolysis is accompanied by a large reduction in particle diameter with a corresponding decrease in surface area during which apo B-100 is responsible for the maintenance of particle integrity. Because the quantity

of apo B-100 in the lipoprotein particle is constant²⁵, as the size decreases the configuration of the protein must change. Only after this compaction and the change in the predominant non-polar lipid from triglyceride to cholesteryl ester does the receptor binding domain of apo B-100 become functional. Some monoclonal antibodies against apo B-100, among them antibodies that block LDL binding to its receptor (4G3 and 3F5), react less well with VLDL than LDL but their immunoreactivity is increased by partial delipidation of the VLDL²⁶. Several apo B-100 epitopes are therefore inaccessible or structurally modified by the configuration adopted by apo B-100 in VLDL but become exposed during the conformational changes that accompany VLDL hydrolysis. Our present data show that these same monoclonal antibodies bind to expressed apo B-100/ β -galactosidase fusion proteins from near the T3/T2 junction (Fig. 3). This region may therefore be involved in the removal of LDL from plasma by the LDL receptor.

Apo B-100 and apo E are the two ligands for the LDL receptor. Detailed studies of apo E established that a region rich in basic amino acids mediates interactions with the LDL receptor^{27–29}. We therefore sought such regions as candidates for the receptor binding domain of apo B-100. There are two basic sequences (residues 3,147–3,157, 3,359–3,367) near the T3/T2 junction³⁰.



Fig. 4 Diagram of the putative receptor-binding domain of apo B-100. Solid blocks, regions enriched in basic amino acids (Fig. 1b, legend). The basic sequence nearer to the carboxy terminal is homologous to the known receptor-binding domain of apo E. Brackets, epitopes of three monoclonal antibodies which block LDL binding to the LDL receptor; arrow, boundary of thrombolytic peptides T3 and T2; S-S, disulphide bridge linking T3 and T2; open circles, N-glycosylation sites; filled circles, sites known to be glycosylated in LDL apo B-100 (Fig. 2a, legend); zigzag, position of the proline-rich cluster of amino acids, which has high amphipathic β -sheet potential.

The second of these sequences has some similarity to the receptor binding domain of apo E (residues 142–150); the consensus is Arg-X-X-Arg-Lys-Arg-X-X-Arg/Lys. These two basic regions are flanked by the epitopes for the receptor blocking monoclonal antibodies (Fig. 4). A study of the disulphide arrangements in apo B-100 (C. H. Spilman *et al.*, unpublished observations) has revealed that the cysteine residues at positions 3,167 and 3,297 (Fig. 2b) are crosslinked. Cys 3,167 is near the T3 basic region and near the epitopes of antibodies 4G3 and 3F5. Cys 3,297 is near the T2 basic region and the epitope of antibody B47. Thus, although the epitopes of these receptor blocking antibodies are distant from one another in the primary structure of apo B-100, they may be very close as a result of the Cys 3,167/Cys 3,297 disulphide linkage. The features of the region of apo B-100 that may form the receptor binding domain are summarized in Fig. 4.

These results have significant implications for the genetics of hyperlipidaemia. LDL has been previously estimated to contain 500K of protein^{31,32}. The M_r predicted from the cDNA sequence indicates that there can be only one molecule of apo B-100 in each LDL particle. Therefore any mutation of the protein that affects lipid metabolism is likely to result in two functionally distinct classes of LDL, one normal and one abnormal in the heterozygous state. This is important as numerous polymorphisms of LDL have been reported in man and other animals and some of these are associated with changes in plasma lipid concentration and with increased risk of atherosclerosis^{33,34}. Two such co-dominantly acting mutations have recently been described^{35,36}. The results presented here make further genetic studies possible and should allow the mechanisms of regulation of VLDL production to be elucidated. In individuals without decreased activity of the LDL receptor, alterations in the structure or rate of production of apo B-100 may be important in determining the level of LDL cholesterol in plasma.

We thank L. Caiati, C. Fortier, E. Ludwig, R. Otter, V. Pierotti and L. Priestley for assistance; G. Bell, T. Belt, M.-L. Chu and L. Rall for supplying cDNA libraries and RNA; S. Young, L. Curtiss and J. Witztum for antibody B47; M. Waterfield for protein sequencing; G. Soundy for computer studies; Graeme Bell and Howard Eder for helpful discussions; and M. Runnicles, L. Sargeant and J. Gilbert for preparation of the manuscript.

Received 3 July; accepted 28 August 1986.

- LaRosa, J. C., Glueck, C. J. & Segal, M. D. *Circulation* **73**, 1–133 (1986).
- Havel, R. J., Goldstein, J. L. & Brown, M. S. in *Metabolic Control and Disease* (eds Bondy, P. K. & Rosenberg, L. E.) 393–494 (Saunders, Philadelphia, 1980).
- Goldstein, J. L. & Brown, M. S. in *The Metabolic Basis of Inherited Disease* (eds Stanbury, J. B. *et al.*) 672–712 (McGraw-Hill, New York, 1983).
- Mahley, R. W., Innerarity, T. L., Rall, S. C. & Weisgraber, K. H. *J. Lipid Res.* **25**, 1277–1294 (1984).
- Kane, J. P. A. *Rev. Physiol.* **45**, 637–650 (1983).
- Breslow, J. L. A. *Rev. Biochem.* **54**, 699–727 (1985).
- Knott, T. J. *et al. Nucleic Acids Res.* **13**, no. 18 (1986).
- Cardin, A. D. *et al. J. biol. Chem.* **259**, 8522–8528 (1984).
- Lee, P. & Breckenridge, W. C. *Can. J. Biochem.* **54**, 42–49 (1976).
- Chou, P. Y. & Fasman, G. D. *A. Rev. Biochem.* **47**, 251–276 (1978).
- Cardin, A. D., Witt, K. R., Barnhart, C. L. & Jackson, R. L. *Biochemistry* **21**, 4503–4511 (1982).
- Segrest, J. P., Jackson, R. L., Morrisett, J. D. & Gotto, A. M. *FEBS Lett.* **38**, 347–353 (1974).
- Kaiser, E. T. & Keady, F. J. *Science* **223**, 249–255 (1984).
- Osterman, D., Mora, R., Keady, F. J., Kaiser, E. T. & Meredith, S. C. *J. Am. chem. Soc.* **106**, 6845–6847 (1984).
- Kyte, J. & Doolittle, R. F. *J. molec. Biol.* **157**, 105–132 (1982).
- Van't Hooft, F. & Havel, R. J. *J. biol. Chem.* **256**, 3963–3968 (1981).
- Südhof, T. C. *et al. Science* **228**, 893–895 (1985).
- Doolittle, R. F. *Trends biochem. Sci.* **10**, 233–237 (1985).
- Freedman, R. B. *Trends biochem. Sci.* **9**, 438–441 (1984).
- Edman, J. C., Ellis, L., Blacher, R. W., Roth, R. A. & Rutter, W. J. *Nature* **317**, 267–270 (1985).
- Yamamoto, T., Bishop, R. W., Brown, M. S., Goldstein, J. L. & Russell, D. W. *Science* **232**, 1230–1237 (1986).
- Christensen, N. J., Rubin, C. E., Cheung, M. C. & Albers, J. J. *J. Lipid Res.* **24**, 1229–1242 (1983).
- Vauhkonen, M., Vitala, J., Parkkinen, J. & Rauvala, H. *Eur. J. Biochem.* **152**, 43–50 (1985).
- Siuta-Mangano, P., Janero, D. R. & Lane, M. D. *J. biol. Chem.* **257**, 11463–11467 (1982).
- Eisenberg, S. & Levy, R. I. *Adv. Lipid Res.* **13**, 1–89 (1975).
- Teng, B. *et al. J. biol. Chem.* **260**, 5067–5072 (1985).
- Innerarity, T. L., Friedlander, E. J., Rall, S. C., Weisgraber, K. H. & Mahley, R. W. *J. biol. Chem.* **258**, 12341–12347 (1983).
- Weisgraber, K. H. *et al. J. biol. Chem.* **258**, 12348–12354 (1983).
- Innerarity, T. L. *et al. J. biol. Chem.* **259**, 7261–7267 (1984).
- Knott, T. J. *et al. Science* **230**, 37–43 (1985).

- Elovson, J., Jacobs, J. C., Schumaker, V. N. & Puppione, D. L. *Biochemistry* **24**, 1569–1578 (1985).
- Deckelbaum, R. J., Shipley, G. G. & Small, D. M. *J. biol. Chem.* **252**, 744–754 (1977).
- Berg, K. & Bearn, A. B. *Clin. Genet.* **1**, 104–120 (1970).
- Rapas, M. & Hasler-Rapas, J. *Fedn Proc.* **72**, 2347–2351 (1985).
- Young, S. G., Bertics, S. J., Curtiss, L. K., Casal, D. C. & Witztum, J. L. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1101–1105 (1986).
- Berg, K. *et al. Proc. natn. Acad. Sci. U.S.A.* (in the press).
- Hardman, D. A., Gustafson, A., Schilling, J. W., Donaldson, V. H. & Kane, J. P. *Biochem. biophys. Res. Commun.* **137**, 821–825 (1986).
- LeBoeuf, R. C. *et al. FEBS Lett.* **170**, 105–108 (1984).
- Belt, K. I., Carroll, M. C. & Porter, R. R. *Cell* **36**, 907–914 (1984).
- Kane, J. P., Hardman, D. A. & Paulus, H. E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2465–2469 (1980).
- Eisenberg, D., Schwarz, E., Komaromy, M. & Wall, R. J. *molec. Biol.* **179**, 125–142 (1984).
- Marcel, Y. L., Hogue, M., Theolis, R. & Milne, R. W. *J. biol. Chem.* **257**, 13165–13168 (1982).
- Young, S. G., Witztum, J. L., Casal, D. C., Curtiss, L. K. & Bernstein, S. *Arteriosclerosis* **6**, 178–188 (1986).
- Ruther, U. & Müller-Hill, B. *EMBO J.* **2**, 1791–1794 (1983).
- Johnston, S., Lee, J. H. & Ray, D. S. *Gene* **34**, 137–145 (1985).
- Burnette, W. N. *Analyt. Biochem.* **112**, 195–203 (1981).

Sequence, structure, receptor-binding domains and internal repeats of human apolipoprotein B-100

Chao-Yuh Yang*, San-Hwan Chen*, Sandra H. Gianturco*, William A. Bradley*, James T. Sparrow*, Masako Tanimura†, Wen-Hsiung Li†, Doris A. Sparrow*, Hans DeLoof‡, Maryvonne Rosseneu‡, Fu-Shin Lee*, Zi-Wei Gu*, Antonio M. Gotto, Jr* & Lawrence Chan*

* Departments of Medicine and Cell Biology, Baylor College of Medicine and The Methodist Hospital, Houston, Texas 77030, USA

† Center for Demographic and Population Genetics, University of Texas, Houston, Texas 77030, USA

‡ Department of Clinical Biochemistry, A.Z. St-Jan, B-8000 Brugge, Belgium

Apolipoprotein (apo) B-100, the major protein component in low density lipoprotein (LDL), is the ligand that binds to the LDL receptor¹. It is important in the metabolism of LDL and elevated plasma levels of LDL-apo B are strongly associated with increased risk of coronary artery disease. Although apo B-100 is of great clinical and biological importance its primary structure has defied chemical elucidation, mainly because of its enormous size, insolubility, and tendency to aggregate². Less than 5% of the apo B-100 sequence has been reported, despite the efforts of many laboratories over the past twenty years. Here we report the complete amino acid sequence of human apo B-100 as deduced by sequence analysis of complementary DNA clones; 2,366 of the 4,536 residues were also confirmed by direct sequencing of apo B-100 tryptic peptides. The distribution of trypsin-accessible and -inaccessible peptides of the protein on LDL is non-random and they can be grouped into 5 hypothetical domains. Of 20 potential N-glycosylation sites identified in the sequence, 13 were found by direct peptide sequencing to be glycosylated, and 4 unglycosylated. Examination of the primary structure of apo B-100 reveals that it contains a large number of long (>70 residues) internal repeats and an even larger number of shorter ones, suggesting that the apo B-100 sequence was derived largely from internal duplications. Finally, using synthetic peptides of a specific region of apo B-100, we have identified a potential LDL receptor-binding domain (residues 3,345–3,381) which can bind to the LDL receptor and suppress 3-hydroxy-3-methyl-glutaryl coenzyme A (HMG-CoA) reductase activities in cultured human fibroblasts.

We have determined the primary structure of human apo B-100 by two complementary approaches: (1) using novel techniques, we determined the sequence of tryptic peptides of apo B-100 and identified over half of the 4,536 amino acid residues in mature apo B-100 directly; and (2) by sequence analysis of 30 overlapping partial cDNA clones, we determined the com-

plete apo B-100 sequence. Because of space limitations the cDNA sequence, which spans over 14 kilobases, will be presented elsewhere³. The complete primary structure of apo B-100 is presented in Figure 1a.

Apo B-100, with a calculated relative molecular mass of 512,937, is one of the largest monomeric proteins known. The sequence is characterized by high hydrophobicity⁴ yielding 0.916 kcal per residue, compared to 0.718, 0.772, 0.806, 0.863, 0.825, 0.838 and 0.752 kcal per residue for apo E, apo A-IV, apo A-I, apo A-II, apo C-I, apo C-II and apo C-III, respectively. Secondary structure analysis⁵ indicates that apo B contains 43%, 21% and 20% α -helical, β -sheet, and random structures, and 16% β -turns, respectively. The high content of β -sheet structure predicted from the primary structure of apo B-100 is consistent with previous observations^{6,7}.

Primary structure analysis of apo B-100 has been greatly hampered by its insolubility in aqueous buffers. Separating the complex peptides generated from proteolytic digestion of a protein of this size is also difficult². In this study we have overcome some of the inherent difficulties and have determined the structure of over 50% of the total apo B-100 sequence. Digestion of LDL by trypsin releases specific peptides from apo B-100 (Figs 1 and 2). When the digestion products were fractionated on a Sephadex G-50 column, the partially-digested LDL eluted in the void volume (T1 of Figure 2a) and the released peptides were partially resolved into multiple peaks (T3 to T7 of Figure 2a). The T1 fraction was delipidated and redigested with trypsin. Repeated fractionations of the redigested T1 peptides and the initially released peptides in T3→T7 resulted in pure peptides. From the sequence of such tryptic peptides (Fig. 2) we identified 2,366 amino acid residues, all of which match the DNA-deduced sequence³. The lack of peptide sequences that fail to match the cDNA-deduced amino acid sequence confirms that the LDL preparation used was pure.

In Fig. 1a, the apo B-100 tryptic peptides sequenced in our laboratory are given in colour. The blue peptides are from the trypsin-accessible regions on LDL and are derived from peptide peaks T3→T7 in Figure 2a. The red peptides represent trypsin-inaccessible peptides and are derived from the void volume peak (T1) in Fig. 2a. The few peptides present in both T1 and T3→T7 are coloured purple. Figure 1b is a linear map of apo B-100 with the differential trypsin-accessibility and Pro and Cys residues highlighted.

The distribution of trypsin-accessible and -inaccessible peptides in apo B-100 is non-random, and there is almost no overlap between these two groups of peptides. The same peptides were accessible to trypsin digestion whether the LDL was reduced and alkylated before or after the enzyme treatment. One can divide apo B-100 into five domains in terms of the clustering of trypsin-accessible peptides (Fig. 1b): domain 1, an NH₂-terminal region (residues 1 to ~1,000) that consists almost exclusively of trypsin-accessible peptides; 2, a shorter region next to it from residue 1,001 to ~1,360 that appears to be transitional between 1 and 3 with alternating accessible and inaccessible peptides; 3, a long domain from residue 1,361 to ~3,160, largely trypsin-inaccessible, with occasional accessible peptides interspersed; 4, a domain from residue 3,161 to ~4,130 of alternating accessible and inaccessible peptides, each 100–200 amino acid residues; and 5, a COOH-terminal domain from residue 4,131 to the COOH-terminus of apo B-100 that consists almost entirely of inaccessible peptides with the exception of the very last decapeptide, which is accessible. The average hydrophobicities of the first four domains (–0.901, –0.800, –0.906, and –0.901 kcal per residue, respectively) were either similar to or somewhat lower than that of the whole molecule (–0.916 kcal per residue). The COOH-terminal domain was significantly more hydrophobic (–1.045 kcal per residue). Its hydrophobic nature may result in a conformation of LDL that renders the region inaccessible to trypsin.

The five hypothetical domains of apo B-100 are highly reproducible from different trypsinization experiments. The trypsin-accessibility of the peptides is related to the conformation of apo B-100 on LDL, which is probably determined by its interaction with the lipid-water environment. It is evident that there is an unusual clustering of Cys residues (coloured orange in Figure 1a, see also Fig. 1b) in the NH₂-terminal region of apo B-100. Twelve of the 25 Cys are found in the NH₂-terminal 500 residues of apo B-100 (14 in the NH₂-terminal domain defined above). These Cys are found entirely on trypsin-accessible peptides (Fig. 1). We speculate that they might form disulphide bridges and help to stabilize apo B-100 structure on LDL. They may also be involved in disulphide linkages with apo(a) on lp(a), lipoprotein particles highly correlated with cardiovascular disease^{8,9}.

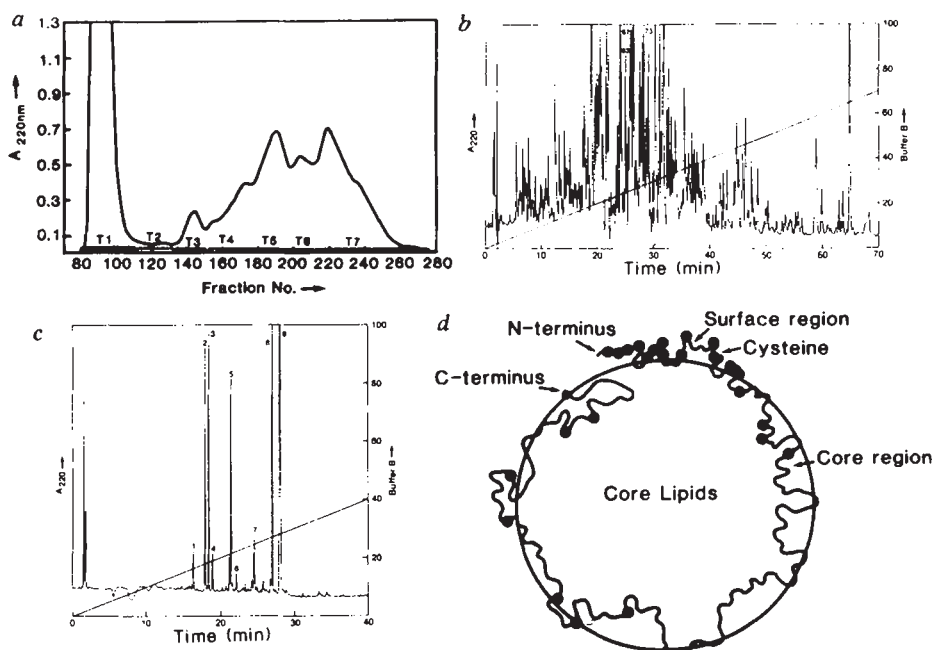
The distribution of the 20 potential N-linked glycosylation sites identified in apo B-100 (coloured green in Figure 1a) is

Table 1 Potential long repeats in Apo B-100, their length (L) and per cent homology (H)

Segment pair	L	H	Segment pair	L	H
Group I			Group II		
938–1249 vs 3231–3542	312	14	467–1007 vs 2611–3151	541	12
1029–1105 vs 3854–3930	77	21	591–667 vs 835–911	77	21
1051–1203 vs 1486–1638	153	17	454–548 vs 1773–1867	95	21
995–1241 vs 2147–2393	247	14	532–681 vs 4298–4447	150	17
2142–2354 vs 4034–4246	213	15	551–532 vs 2846–2927	82	21
4057–4129 vs 4316–4388	73	26	748–827 vs 2768–2847	80	20
1432–1788 vs 3124–3480	357	12	811–882 vs 1286–1357	72	21
1544–1620 vs 1656–1732	77	23	2670–2752 vs 3249–3331	83	23
1444–1534 vs 3456–3546	91	20	2826–2903 vs 3070–3147	78	22
1492–1646 vs 1758–1912	155	17	2646–2842 vs 4336–4532	197	15
1498–1579 vs 3976–4057	82	20	63–142 vs 4335–4414	80	20
1676–1749 vs 4444–4517	74	20			
1761–1852 vs 1906–1997	92	24	Group III		
1882–2000 vs 2988–3106	119	23	32–112 vs 3861–3941	81	21
1500–1672 vs 3372–3544	173	16	2418–2489 vs 3861–3941	81	20
2795–2975 vs 3366–3546	181	16	2388–2470 vs 4221–4303	83	22
1332–1644 vs 3333–3645	313	13	3840–3917 vs 4214–4291	78	21
2927–3105 vs 3456–3634	179	16	2505–2594 vs 4201–4290	90	20
1293–1486 vs 1677–1870	194	16	2352–2441 vs 4201–4290	90	21
1279–1358 vs 2875–2954	80	20			

Homology is defined as the proportion of identical amino acids between the two segments compared. Most of the homology values in the table are comparable to the values observed between the 22-residue repeats in human apo A-I¹⁵; for example, in human apo A-I, the homology values between the repeat 99–120 and the repeats 44–65, 66–87, 121–142, 143–164, 165–186 and 220–241 are 14%, 27%, 14%, 9%, 18% and 27%, respectively.

Fig. 2 Purification and sequencing of peptides released from trypsin-treated LDL, and a model of apo B-100 configuration on LDL. Reduced alkylated LDL (200 mg; 40 mg protein) in 20 ml of 0.1 M NH_4HCO_3 , pH 8.0 was subjected to digestion with trypsin (enzyme: substrate = 1:40). The peptide mixture was fractionated on a 2.6×200 cm Sephadex G-50 column. **a**, Absorbance profile of the Sephadex G-50 eluents. Fraction T1 consists of trypsin-inaccessible peptides associated with lipids, T2 of very small amounts of peptide material with trypsin and T3-T7 contain the released peptides. T1 was delipidated in ether:ethanol (3:1 v/v) three times, suspended in 24 ml 0.1 M NH_4HCO_3 , pH 8, and trypsin (0.48 mg) was added. After 24 h the reaction mixture became clear and was loaded on a reverse phase Vydac C_{18} column (10×250 mm) using a trifluoroacetic acid (TFA) buffer system. **b**, Separation of T1 peptides after second tryptic digestion. Buffer A = 0.1% TFA in H_2O ; buffer B = 0.08% TFA; 95% acetonitrile, 5% H_2O . A straight concentration gradient from 0 to 70% buffer B in 70 min was used. The individual peaks were subjected to NH_2 -terminal analysis and sequencing by the modified Edman degradation method. A few of the peptides at this stage are pure. Most were found to be mixtures and were repurified by the rechromatography system of Yang *et al.*²⁶ **c**, Rechromatography of peak no. 63. The separation was performed on a Shandon Hypersil ODS 5μ reverse phase column (4.6×250 mm) with sodium phosphate buffer system. Buffer A = 0.005 M $\text{NaH}_2\text{PO}_4/\text{Na}_2\text{HPO}_4$, pH 6.0; Buffer B = 10% A + 90% CH_3CN . A straight concentration gradient from 0-40% in 40 mins was used. NH_2 -terminal and amino acid analysis of aliquots of these rechromatographed fractions indicated that they were pure. The sequences of all 9 peptides (which matched the DNA-deduced sequence) are as follows: (1) EELCTMFIR (position 4,187-4,195); (2) EYSGTIASEANTYLNSK (3,481-3,497); (3) YEVDQIQVLMMDK (2,321-2,333); (4) NTFLLSCDGLSR (1,389-1,400); (5) EYSGTIASEANTY (3,481-3,493); (6) LEIQSQVDSQVGHVSLTAK (2,959-2,978); (7) NTLELSNGVIVK (2,838-2,849); (8) EVGTVLSSQVYSK (4,196-4,207); (9) SEILAHWSPAK (1,108-1,118). Peptides released initially and separated on the Sephadex G-50 column (T3-T7) were rechromatographed as **b** above. Peptides that were not resolved were rechromatographed on the Shandon Hypersil reverse phase column (as above) to give pure peptides. **d**, Hypothetical model of apo B-100 structure on LDL. For simplicity, the trypsin-accessible regions are shown on the surface, and the trypsin-inaccessible regions inside the core of the LDL particle (not to scale). The surface-core orientation is presented to allow easy visualization of trypsin-accessibility. It does not imply that the different parts of apo B-100 are physically on the surface or buried inside the lipoprotein particle.



also non-random. By direct sequence analysis, 13 sites (residues 956, 1,341, 1,350, 1,496, 2,955, 3,074, 3,197, 3,309, 3,331, 3,384, 3,438, 3,868 and 4,210) were found to be glycosylated and 4 sites (residues 7, 429, 2,212 and 2,533) were not. A cluster of 6 potential sites in the region that includes residues 3,050-3,450, all of which contain carbohydrate is of note. This region of apo B-100 is in close proximity to a potential receptor-binding domain identified in our laboratory (see below).

A hypothetical model of the conformation of apo B-100 on LDL is given in Fig. 2d (compare with Fig. 1b). Although it can be seen that susceptibility to digestion by trypsin does not necessarily imply a surface location for a peptide, or vice versa, we believe that this model, based on information in Figs 1 and 2a-c, represents a testable hypothesis of apo B-100 structure that will be of help in future experiments to elucidate the spatial orientation of apo B-100 on LDL. It will undoubtedly have to be modified as our understanding of apo B-100 structure improves.

When apo B-100 was digested with trypsin the amino-acid composition of the remaining parts of the protein showed little difference from that of the undigested protein^{10,11}. Furthermore, the amino-acid composition of the COOH-terminal fifth of apo B-100 (836 residues) differs only slightly from that of the whole molecule¹². These observations suggest that apo B-100 contains internal repeats. This is interesting because it might explain how such an exceptionally large protein has evolved and because internal repeats have been found in all the other apolipoproteins¹³⁻¹⁵. To investigate this possibility, we have analysed the apo B-100 sequence using the dot matrix method¹⁶ (Fig. 3b). For comparison, we also show a dot matrix for apo A-I (Fig. 3a). It is known that apo A-I contains many consecutive repeats of a 22-residue segment, and these appear in Fig. 3a as lines parallel to one another. In contrast, apo B-100 apparently contains not only a great many short repeats, but also numerous long repeats.

As Fig. 3b is difficult to decipher visually, we present a partial list of the potential long repeats in apo B-100 in Table 1. The potential repeats listed are divided into three groups such that segments within a group are more similar to one another than to segments in other groups. In group I, there are three potential repeats with lengths greater than 300 residues: 938-1,249 vs 3,231-3,542 (312 residues and 14% homology), 1,432-1,788 vs 3,124-3,480 (357 residues and 12% homology), and 1,332-1,644 vs 3,333-3,645 (313 residues and 13% homology). These also have smaller regions of homology with other regions, for example, segment 938-1,249, contains a subsegment (1,029-1,105) which has 21% homology with the segment 3,854-3,930. The segment 1,432-1,788, which contains the repeat pair 1,544-1,620 and 1,656-1,732, also contains many subsegments that have significant homologies to other segments. This is also true for the segment 1,332-1,644. This last segment overlaps to a large extent with the segment 1,432-1,788, which contains a segment (1,486-1,638) that has 17% homology with a subsegment (1,051-1,203) of the segment 938-1,249. Therefore, all the segments in group I might be evolutionarily related to one another. The same might be true for the segments in group II and for those in group III.

The potential repeats listed in Table 1 were chosen either because they have a length ≥ 70 residues and a homology $\geq 20\%$ or because they have a length ≥ 150 residues and a homology $\geq 12\%$. It is interesting to note that these potential repeats almost completely cover the entire apo B-100 sequence. Since Fig. 3b shows many more potential repeats than those listed in Table 1, we conclude that the apo B-100 sequence was derived almost entirely from internal duplications.

Apo B-100 on LDL is the ligand recognized by the LDL receptor. Brown and Goldstein¹ have shown that the LDL receptor contains a domain at the NH_2 -terminal portion of the receptor which consists of Cys-rich repeating units. A cluster of negatively charged amino acids occurs near the C-terminus of

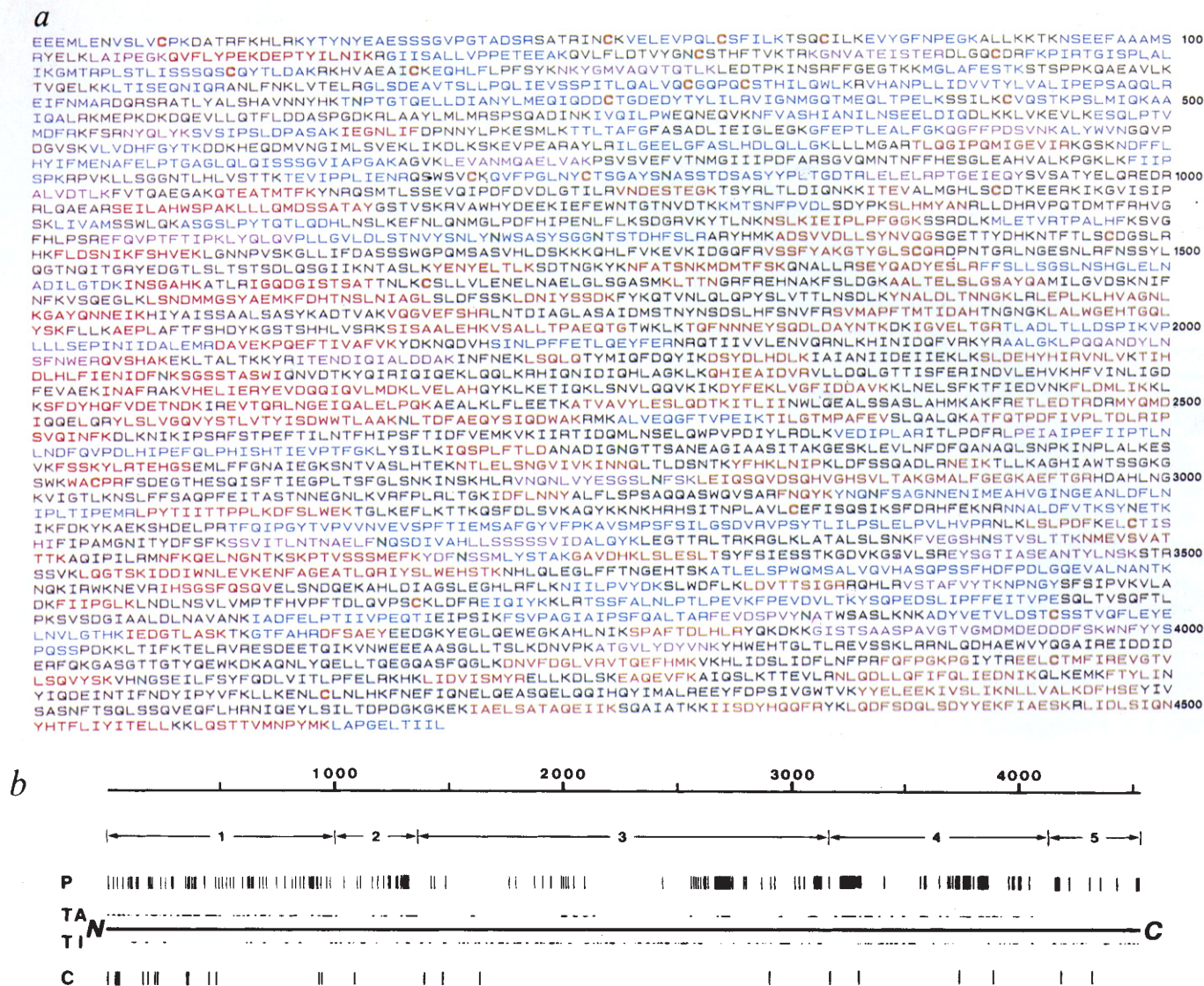


Fig. 1 **a**, Primary structure of apo B-100 and distribution of trypsin-accessible versus trypsin-inaccessible peptides. The complete sequence of apo B-100 is based on the combined data from the cDNA-deduced amino acid sequence³ and sequences from the tryptic peptides. The red peptides are those that are protected from trypsin cleavage during the initial tryptic digestion, and are released from the first peak (T1) on the Sephadex G-50 column (Fig. 2a) on delipidation and retypsinization. The blue peptides are the trypsin-accessible peptides (T3 → T7; Fig. 2a). The orange Cs are the Cys residues. Green N, potential N-glycosylation sites (Asn-X-Thr/Ser). The amino acids are shown in single letter codes: A = Ala, C = Cys, D = Asp, E = Glu, F = Phe, G = Gly, H = His, I = Ile, K = Lys, L = Leu, M = Met, N = Asn, P = Pro, Q = Gln, R = Arg, S = Ser, T = Thr, V = Val, W = Trp, Y = Tyr. **b**, Linear map of apo B-100. The heavy line (N—C) represents the linear sequence with its NH₂- and COOH-termini. The trypsin-accessible (TA) domains are marked by a line above, and the inaccessible (TI) domains by a line below the linear sequence. Prolines (P) are denoted by vertical bars above, and cysteines (C) by bars below the sequence. The five domains identified in the text are also marked. Scale, number of amino-acid residues.

each repeat. Brown and Goldstein speculate that these charged residues are complementary to a cluster of positively charged residues on the receptor-binding domain(s) of apo B-100. There is a similar cluster in apo E^{17,18}. While our study was in progress, Knott *et al.*¹⁹ identified, on the basis of a partial cDNA sequence, a region of apo B that shows significant homology to apo E (140–150), the putative LDL receptor binding domain. They speculate that this region may mediate the binding of apo B to the LDL receptor. We have chemically synthesized this region of apo B, which corresponds to residues 3,345–3,381 of our sequence and, as a control, a second region of apo B (4,154–4,189) which binds lipid but is not homologous to apo E (140–150) receptor binding domain. We tested these synthetic apo B peptides for binding to the LDL receptor of cultured human fibroblasts. Apo B 3,345–3,381, but not apo B 4,154–4,189, can indeed mediate binding and uptake by the LDL receptor pathway.

To test whether apo B synthetic peptides bind to the LDL receptor, we used our previous observation that trypsinization of hypertriglyceridemic (t-HTG)-very low density lipoproteins (VLDL) abolishes their ability to bind to the LDL receptor and, therefore, their ability to suppress HMG-CoA reductase activity in normal human fibroblasts. This is an exquisitely sensitive intracellular endpoint that requires LDL receptor-mediated binding, uptake, and degradation of a particle^{20–23}. We chose this test system because of its sensitivity and immunity from the complications inherent in interpreting direct binding studies performed with lipoproteins containing readily transferable, soluble peptides^{20,24,25}.

Figure 4 illustrates the effects of t-HTG-VLDL in the presence and absence of apo B synthetic peptides on HMG-CoA reductase activity in normal fibroblasts. t-HTG-VLDL do not suppress reductase activity because they do not contain intact apo E^{20,24,25}. The synthetic apo B peptide 4,154–4,189 failed to confer

upon the re-isolated trypsinized VLDL the ability to bind to LDL receptor and effectively suppress HMG-CoA reductase activity. By contrast, the putative receptor binding region apo B 3,345-3,381, when bound to t-HTG-VLDL, restored the ability of these particles to suppress HMG-CoA reductase activity in an efficient, saturable manner. This is analogous to our previously published studies showing that incorporation of intact apo E into t-HTG-VLDL (or into thrombin-inactivated VLDL, or into normal VLDL) confers upon these particles the ability to bind to the LDL receptor²⁰⁻²³. Two different preparations of the apo B synthetic peptides and two different trypsinized VLDL preparations in two separate cell experiments gave similar results. Thus the synthetic peptide apo B 3,345-3,381, but not apo B 4,154-4,189, contains a domain that can bind to the LDL receptor of cultured human fibroblasts. The data support the prediction that the receptor binding domain on LDL contains a positively charged amino-acid cluster similar to that found in apo E^{17,18}. We have not, however, excluded the possibility of additional receptor binding domains of apo B-100.

Thus we have determined the primary structure and defined some structure-function relationships of apo B-100. We have mapped the trypsin-accessible and -inaccessible domains of the protein on LDL, identified numerous long internal repeats in apo B-100, and defined a putative LDL receptor binding domain of apo B-100. The observations reported here form a basis for future investigations of the functional domains of apo B-100 in lipid transport, antibody recognition, receptor interactions and atherosclerosis.

We thank Drs C. B. Lawrence and D. A. Goldman for assistance with computer analysis, Joseph Hagman, Shan-Zhan Ye, A. Lin, S. L. Hwang, and F. Brown for technical assistance, and

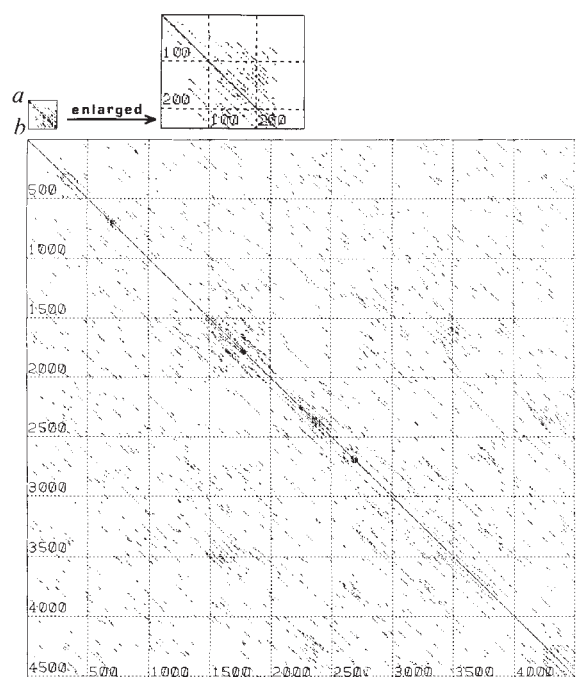


Fig. 3 *a*, Dot matrix for the human apo A-I mature peptide and *b*, dot matrix for the human apo B-100 mature peptide. Algorithm by Lawrence *et al.*¹⁶. Each region of similarity with a minimum segment length of 20 residues and a minimum similarity score of 4.5 is displayed as a linear set of dots; each dot represents a perfect match of amino acids. The score is defined as the number of standard deviations by which the number of matches between the two segments compared exceeds chance¹⁶. It is known that apo A-I contains many consecutive repeats of a 22-residue segment¹³⁻¹⁵; these are seen as lines parallel to one another in matrix *a*. Compared with matrix *a*, it is clear that matrix *b* contains not only a large number of short lines, but also many lines much longer than 22 residues.

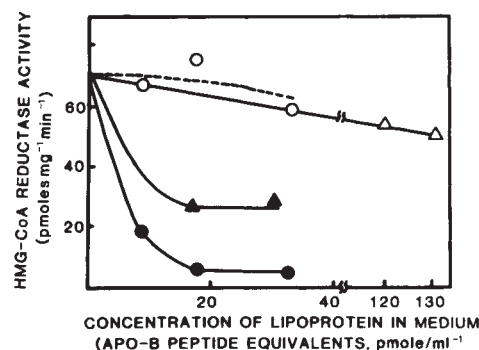


Fig. 4 Apo B (3,345-3,381; that is, SVIDALQYKLEGTRTLTR-KRGLKLATALSLSNKFVEG) reincorporation restores LDL receptor binding activity to trypsin-inactivated HTG-VLDL S_f 100-400 (where S_f 100-400 is the population of lipoproteins identified with the flotation rate 100-400 negative Svedberg units; HTG-VLDL₁). Cultured human skin fibroblasts were grown to approximately 75% confluency in complete medium containing 10% fetal calf serum, washed, and placed on 2 ml of medium containing 5% lipoprotein deficient serum for 24 h²⁰⁻²³. Indicated amounts of lipoproteins or reconstituted lipoproteins, in 0.2 ml, were added to duplicate dishes for 16 h before cells were washed, harvested, and assayed for HMG-CoA reductase activity²⁰. Each data point is the average of duplicate determinations. The average variation between dishes was <5%. Hypertriglyceridemic VLDL S_f 100-400, HTG-VLDL₁, were isolated by cumulative flotation from the plasma of a patient with type 4 hyperlipoproteinemia²⁰⁻²³. Trypsin-inactivated HTG-VLDL₁ (450 μg of protein) were incubated in the presence and absence of ¹²⁵I-labelled peptide (100 μg) for 2 h at 37 °C, then for 16 h at 4 °C; they were then reisolated by flotation as described previously²⁰ before testing in fibroblasts. Trypsin-treated HTG-VLDL₁ (○); trypsin-treated HTG-VLDL₁ + apo B (4,154-4,189) (Δ); trypsin-treated HTG-VLDL₁ + apo B (3,345-3,381) (▲); and LDL (●).

S. Saltalamacchia for secretarial assistance. S.H.G. and J.T.S. are Established Investigators of the American Heart Association. We thank Dr Larry Mathis and Mr Albert Alkek for their support and interest. Supported by National Institutes of Health grant HL-27341 for a Specialized Center of Research in Atherosclerosis, and GM-30998 (W.H.L.), and by grants from the March of Dimes Birth Defects Foundation, The Methodist Hospital, Texas Eastern Corporation, and the Pauline Sterne Wolff Memorial Foundation.

Received 3 July; accepted 28 August 1986.

1. Brown, M. S. & Goldstein, J. L. *Science* **232**, 34-47 (1986).
2. Bradley, W. A., Rhode, M. F. & Gotto, A. M. Jr. *Ann. NY Acad. Sci.* **348**, 87-103 (1980).
3. Chen, S. W. *et al. J. Biol. Chem.* **261**, 12918-12921 (1986).
4. Bull, H. B. & Breese, K. *Arch. Biochem. Biophys.* **161**, 665-670 (1974).
5. Garnier, J., Osguthorpe, D. J. & Robson, B. *J. molec. Biol.* **120**, 97-120 (1978).
6. Gotto, A. M., Jr., Levy, R. I. & Frederickson, D. S. *Proc. natn. Acad. Sci. U.S.A.* **60**, 1436-1441 (1968).
7. Chen, G. C. & Kane, J. P. *Biochemistry* **14**, 3357-3362 (1975).
8. Berg, K. *Acta path. microbiol. Scand.* **59**, Suppl. 166-167, 369-382 (1963).
9. Albers, J. J., Adolphson, J. L. & Hazzard, W. R. *J. Lipid Res.* **18**, 331-338 (1977).
10. Margolis, S. & Langdon, R. G. *J. biol. Chem.* **241**, 485-493 (1966).
11. Chapman, M. J., Goldstein, S. & Mills, G. L. *Eur. J. Biochem.* **87**, 475-488 (1978).
12. Wei, C.-F. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 7265-7269 (1985).
13. Karathanasis, S. K., Zannis, V. I. & Breslow, J. L. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6147-6151 (1983).
14. Boguski, M. S., Elshourbagy, N., Taylor, J. M. & Gordon, J. I. *Proc. natn. Acad. Sci. U.S.A.* **82**, 992-996 (1985).
15. Luo, C. C., Li, W. H., Moore, M. N. & Chan, L. *J. molec. Biol.* **187**, 325-340 (1986).
16. Lawrence, C. B., Goldman, D. A. & Hood, R. T. *Bull. Math. Biol.* (in the press).
17. Mahley, R. W. & Innerarity, T. L. *Biochim. biophys. Acta* **737**, 197-222 (1983).
18. Innerarity, T. L., Weisgraber, K. H., Arnold, K. S., Rall, S. C. Jr & Mahley, R. W. *J. biol. Chem.* **259**, 7261-7267 (1984).
19. Knott, T. J. *et al. Science* **230**, 37-43 (1985).
20. Gianturco, S. H. & Bradley, W. A. *Meth. Enzym.* **128**, 319-344 (1986).
21. Gianturco, S. H. *et al. J. biol. Chem.* **258**, 4526-4533 (1983).
22. Bradley, W. A. *et al. J. biol. Chem.* **259**, 14728-14735 (1984).
23. Bradley, W. A. & Gianturco, S. H. *J. Lipid Res.* **27**, 40-48 (1986).
24. Havekes, L., Schouten, D., van Hinsbergh, V. & de Wit, E. *Biochem. biophys. Res. Commun.* **122**, 785-790 (1984).
25. Fidge, N. H. & Nestel, P. J. *J. biol. Chem.* **260**, 3570-3575 (1985).
26. Yang, C. Y., Pauly, E., Kratzin, H. & Hilschmann, N. *Hoppe-Seyler's Z. Physiol. Chem.* **362**, 1131-1146 (1981).

Osteoclasts and haematopoietic stem cells in developing human bones

THE stimulating paper by Scheven *et al.*¹ concluded that murine osteoclast-forming capacity rises with increasing haematopoietic stem cell purity when cocultured with preosteoclast-free embryonic long bones. However, a growth-factor-like effect of introduced stem cells should be considered.

In 1982 we called attention to the fact that the appearance of osteoclasts in May-Grünwald-Giemsa-stained smears and serial sections from different bones in the course of human haematopoietic development precedes the onset of haematopoiesis in each bone by about two weeks². This accords with the development of the haematopoietic stroma. So if osteoclasts are derived from haematopoietic stem cells their presence in the marrow could be regarded as a very early 'footprint' of these otherwise undetectable stem cells.

Osteoclasts, formerly believed to be of the monocyte/macrophage lineage could be derived from a marrow stem cell unrelated to the stem cell for other blood cells. There could also still be some doubt about the nature of giant cells appearing during long-term bone marrow culture, that is, in long-term culture of newborn rabbit bone marrow giant cells have been characterized as macrophage polykaryons rather than osteoclastic giant cells³.

E. KELEMEN

First Department of Medicine,
Semmelweis University,
Korányi u 2/A,
1083 Budapest, Hungary

1. Scheven, B. A. A., Visser, J. W. M. & Nijweide, P. J. *Nature* **321**, 79-81 (1986).
2. Kelemen, E. & Calvo, W. in *The Human Bone Marrow* (eds Trubowitz, S. & Davis, S.) Vol. 1, 3-41 (CRC, Boca Raton, Florida, 1982).
3. Horton, M. A., Rimmer, E. F. & Chambers, T. J. *J. Bone Mineral Res.* **1**, 5-14 (1986).

SCHEVEN AND NIJWEIDE REPLY—Because the haematopoietic cell populations used in our study were not 100% pure, we cannot completely exclude the possibility that other stem cells, not identical to the haematopoietic stem cell, were co-purified and gave rise to osteoclast formation. However, the fact that interleukin-3 pretreatment of the stem cell preparations stimulated osteoclast development, as well as formation of granulocytes and macrophages, strengthen the possibility that the haematopoietic stem cell is also the progenitor for osteoclasts.

The first appearance of osteoclasts in developing embryonic bones should not be regarded as a footprint of the presence of haematopoietic stem cells. Osteoclasts differentiate in the cartilaginous models of long bones either from committed

haematogeneous progenitor cells seeded in the perichondrium or from haematopoietic multipotent progenitors committed to differentiate into osteoclasts under the influence of the long bone model. Osteoclasts excavate a marrow cavity in the calcified hypertrophic cartilage area, creating room for stroma formation. Osteoclasts and stromal cells are together responsible for the establishment of a microenvironment for homing and maintenance of haematopoietic stem cells, which are transported via the blood stream from the early haematopoietic organ (liver). Considering this sequence of events it is not surprising that the osteoclasts appear before the onset of bone marrow haematopoiesis.

Finally, we are convinced that the multinucleated cells formed during co-culture with periosteal metatarsal bones are true osteoclasts and not macrophage polykaryons. The multinucleated cells have morphological, enzyme-histochemical (tartrate-resistant acid phosphatase) and functional (resorption) characteristics of osteoclasts. Furthermore, in several non-published experiments we have used quail bone marrow as exogenous source of osteoclast progenitors. The multinucleated cells developed during co-culture with fetal mouse long bones could be recognized by their specific chromatin organization as quail cells, and by their reaction with osteoclast-specific monoclonal antibodies¹ as osteoclasts.

BEN A. A. SCHEVEN
PETER J. NIJWEIDE

Laboratory of Cell Biology
and Histology,
University of Leiden,
Rijnsburgerweg 10,
3233 AA Leiden, The Netherlands

1. Nijweide, P. J. *et al. Histochemistry* **83**, 315-324 (1985).

Interactions of N-CAM with heparin-like molecules

COLE *ET AL.*¹ tried to answer the question of whether the binding of heparan sulphate to the neural cell adhesion molecule, N-CAM, is also required for cell adhesion. They showed that the binding of retinal probe cells to retinal cell monolayers was inhibited by heparin, a substance similar to heparan sulphate; but not by chondroitin sulphate. Monoclonal antibodies that recognized two different domains on N-CAM, the homophilic and heparin-binding domains, inhibited cell adhesion. The heparin-binding domain isolated from N-CAM by selective proteolysis also inhibited cell-cell adhesion when bound to the probe cells.

Heparan sulphate is one of the major constituents of the extracellular matrix, is found on cell surfaces, and throughout the

mammalian body either in its free form or bound covalently or electrostatically by ionic forces to proteins. Heparin is found in small amounts in the mammalian system and its biosynthesis, functions and degradation involve enzyme systems different from those involved in the metabolism of heparan sulphate. The distinctiveness of the two metabolic systems is expressed in human pathology as, for example, Sanfilippo disease type A, characterized by the lack of heparan sulphamidase. Only changes in heparan sulphate metabolism can be detected; heparin metabolism is unaffected. This is cited to show the biological and biochemical differences between those two compounds. The structural differences between the two acid glycosaminoglycans are well documented. The major difference is the degree of sulphation and the length of chains. As the degree of sulphation is responsible for the binding mechanisms of this highly negatively charged polyanion, heparin should have been used rather as a control substance in addition to the use of heparan sulphate in the protein interaction studies.

The use of heparitinase, which like heparan sulphate, is available, could have helped or could be suggested for confirmation of the data obtained.

Chondroitin sulphate is not a good control in our opinion as it is completely different from heparan sulphate and heparin: the sulphate in this compound is bound to the polysaccharide skeleton by O-sulphation and not, as in heparan sulphate and heparin by N-sulphation. Thus, even if chain length and degree of sulphation of the glycosaminoglycan were comparable, the unknown charge distribution of 'chondroitin sulphate mixed isomers'¹ could not be compared. The reactivity of heparan sulphate (and heparin) with proteins is enormous. This has been shown by several authors. Gelman and coworkers studied the glycosaminoglycan-protein interaction using circular dichroism² and Doyle and associates published the interaction between collagen and acid glycosaminoglycans indicating that sulphated glycosaminoglycans were binding (nonspecifically) to clustered, basic, positively charged amino-acid residues³.

Laminin, an intrinsic glomerular basement membrane protein which is involved in cell-basement membrane interactions, and fibronectin interact with heparan sulphate and heparin^{4,5}. So does collagen type IV and the native glomerular basement membrane, containing covalently linked heparan sulphate-proteoglycan⁶. Thus, heparan sulphate seems to be a highly reactive substance resembling other compounds (such as laminin, actin and fibronectin) with 'glue-activity'.

In our hands this acid-glycosaminoglycan reacts with commercially available polylysine as a marker substance for

'clustered positively charged groups'. Antibodies bound to the heparan sulphate-binding domain can change the charge of this peptide or can inhibit it by steric hindrance thus interfering with the heparan sulphate binding.

GERT LUBEC

Université Rene Descartes,
Laboratoire de Biochimie,
45 rue des Saints Peres,
F 75270 Paris, France

1. Cole, G. J., Loewy, A. & Glaser, L. *Nature* **320**, 445-447 (1986).
2. Gelman, R. A. *et al. Biochim. biophys. Acta* **427**, 492-496 (1976).
3. Doyle, B. B. *et al. J. molec. Biol.* **91**, 79-99 (1974).
4. Lubec, G. *Renal Physiol.* **9**, 9-17 (1986).
5. Lubec, G. *Ped. Pedol.* **17**, 591-596 (1982).
6. Parthasarathy, N. & Spiro, R. G. *J. biol. Chem.* **256**, 507-512 (1981).

COLE ET AL. REPLY—Heparin was used as a structural analogue of heparan sulphate in our work because of the similarity in structure between the two molecules. Previous studies involving laminin or fibronectin used heparin-agarose to demonstrate that these molecules possessed domains for interaction with heparin^{1,2}. However, it is now well documented that the actual cell ligand that interacts with these molecules is heparan sulphate-proteoglycan. Thus, heparin was used as an analogue of heparan sulphate. In our studies³ both heparin and heparan sulphate, which are structurally related, inhibit cell-cell adhesion and therefore heparin would not serve as an adequate control. Chondroitin sulphate was used as a control to demonstrate that a specificity existed for the glycosaminoglycan-binding domain. We have also shown that dermatan sulphate or hyaluronic acid do not inhibit heparin binding to N-CAM (unpublished observations). If nonspecific binding was occurring to a cluster of basic amino acids, it would be assumed that chondroitin sulphate (or other glycosaminoglycans) would also bind non-specifically. Additionally, other glycosaminoglycans (chondroitin sulphate and hyaluronic acid) also interact with fibronectin, and these interactions occur at domains distinct from the heparin-binding region¹. It is thus apparent that these glycosaminoglycans are interacting specifically with these adhesive molecules, and that chondroitin sulphate is an appropriate control.

With regard to Lubec's comment about the type of sulphation on chondroitin sulphate and heparin/heparan sulphate, he states that chondroitin sulphate is *O*-sulphated and heparin/heparan sulphate is *N*-sulphated. He therefore concludes that chondroitin sulphate is not an appropriate control because the *O*-sulphation may contribute to a different charge distribution than observed with heparin/heparan sulphate. It should be noted that heparin is also *O*-sulphated, and in fact a 3-*O*-sulphate group on a

pentasaccharide fragment derived from heparin is required for the ability of heparin to inhibit the proliferation of vascular smooth muscle cells⁴. Our laboratory has used fractionated heparin molecules to obtain similar results, as heparin molecules that inhibit smooth muscle cell proliferation also bind to N-CAM (manuscript in preparation). Therefore, because both chondroitin sulphate and heparin contain *O*-sulphate groups, but only heparin binds to N-CAM, it appears that the heparin-binding domain of N-CAM exhibits glycosaminoglycan specificity.

GREGORY J. COLE
ARLEEN LOEWY
LUIS GLASER

Department of Biological Chemistry,
Washington University School
of Medicine,
St Louis, Missouri 63110, USA

1. Yamada, K. M. *A. Rev. Biochem.* **52**, 761-799 (1983).
2. Hynes, R. O. & Yamada, K. M. *J. Cell Biol.* **95**, 369-377 (1982).
3. Cole, G. J., Loewy, A. & Glaser, L. *Nature* **320**, 445-447 (1986).
4. Castellot, J. J. *et al. J. Cell Biol.* **102**, 1979-1984 (1986).

Evolutionary relationships of human populations

WAINSCOT ET AL.¹ state that the phylogenetic tree of human races based on blood groups has its primary division between the mongoloid and the caucasoid-negroid groups, in contrast to the one they derived from nuclear DNA polymorphisms. This is a misreading of their quoted source, a paper by Professor L. L. Cavalli-Sforza and me² in which we constructed a tree for 15 populations from just 5 blood-group systems. Apart from the fact that this tree was computed 23 years ago and should thus be remembered (if at all) for the originality of the method rather than the definitiveness of the result, we wrote "It should be noted that the origin of the tree, which we have set near the centre of the system, is arbitrary, but the splits and the computed segment lengths would be unchanged whatever the origin".

The reason for this is that we used our 'method of minimum evolution'³ (now often called the 'principle of parsimony', although that phrase properly belongs to the classical geometrical notion that figures are to be constructed with ruler and compass only⁴) which does not allow an origin to be inferred as it does not explicitly allow for a time dimension. Indeed, on another occasion on which we presented the same tree⁵ we used a different arbitrary origin.

In fact our published tree² shows that the longest segment by far is between the African group (Ethiopian, Bantu and Ghanaian) and the rest, a feature amply confirmed by the associated analysis into two and three principal components. To

paraphrase the summary of Wainscoat *et al.*¹, "genetic distance analysis based on these blood-group polymorphisms indicates a major division of human populations into an African and a Eurasian group".

A. W. F. EDWARDS

Department of Community Medicine,
University of Cambridge,
Gresham Road,
Cambridge CB1 2ES, UK

1. Wainscoat, J. S. *et al. Nature* **319**, 491-493 (1986).
2. Cavalli-Sforza, L. L. & Edwards, A. W. F. *Genet. Today* (Pergamon, Oxford, 1964).
3. Edwards, A. W. F. & Cavalli-Sforza, L. L. *Heredity* **18**, 553 (1963).
4. Boyer, C. B. *A History of Mathematics* (Wiley, New York, 1968).
5. Cavalli-Sforza, L. L., Barrai, I. & Edwards, A. W. F. *Cold Spring Harbor Symp. quant. Biol.* **29**, 9-20 (1964).

WAINSCOT ET AL. REPLY—We accept Dr Edwards's interpretation of the classic paper¹ by himself and Professor Cavalli-Sforza in that, although their phylogenetic tree (their Fig. 5) does have the first split between the mongoloid and caucasoid-negroid groups, this division is arbitrary. However, much more data both on blood groups and protein loci have been obtained over the past 20 years and this has recently been comprehensively analysed by Nei and Roychoudhury². Interestingly, there is still a discrepancy between the evolutionary analyses based on blood groups and protein loci as stated in our paper³. Thus, the total blood group data now available are inconsistent with Dr Edwards's suggested paraphrase of our summary.

J. S. WAINSCOT
A. V. S. HILL
S. L. THEIN
J. B. CLEGG

Department of Haematology
and MRC Molecular
Haematology Unit,
John Radcliffe Hospital,
Headington, Oxford OX3 9DU, UK

1. Cavalli-Sforza, L. L. & Edwards, A. W. F. *Genetics Today* (Pergamon, Oxford, 1964).
2. Nei, M. & Roychoudhury, A. K. *Evol. Biol.* (Plenum, New York, 1982).
3. Wainscoat, J. S. *et al. Nature* **319**, 491-493 (1986).

Matters Arising

Matters Arising is meant as a vehicle for comment and discussion about papers that appear in *Nature*. The originator of a Matters Arising contribution should initially send his manuscript to the author of the original paper and both parties should, wherever possible, agree on what is to be submitted. Neither contribution nor reply (if one is necessary) should be longer than 500 words and the briefest of replies, to the effect that a point is taken, should be considered.